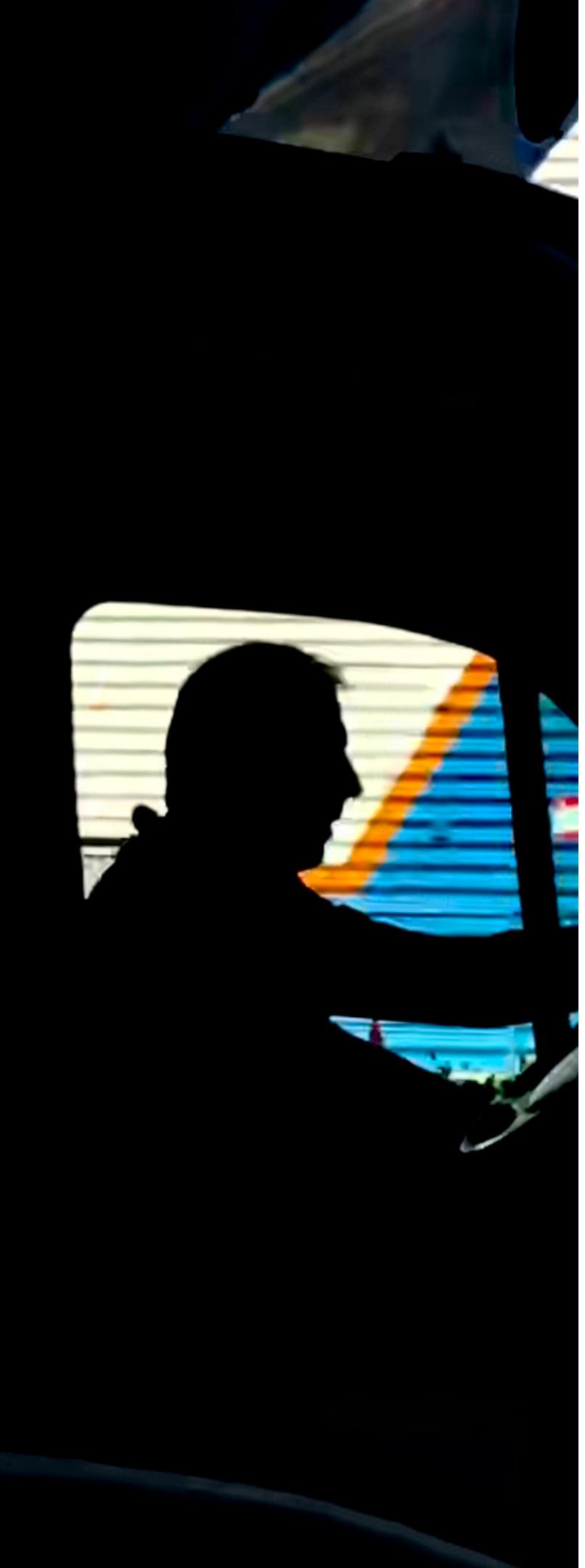


Pakistan's Largest Ecommerce Dataset



Ilham Nashrullah H / JCDS 1904 / Purwadhika





Content

- 1 Pendahuluan**
- 2 Struktur Data**
- 3 Data Preparation**
- 4 Statistika**
- 5 Pembahasan**



1

Pendahuluan

1.1 Deskripsi Dataset

Dataset ini merupakan kumpulan data pesanan retail e-commerce terbesar dari Pakistan.

Dataset ini berisi setengah juta catatan transaksi dari Maret 2016 hingga Agustus 2018. Data dikumpulkan dari berbagai pedagang e-commerce sebagai bagian dari studi penelitian.

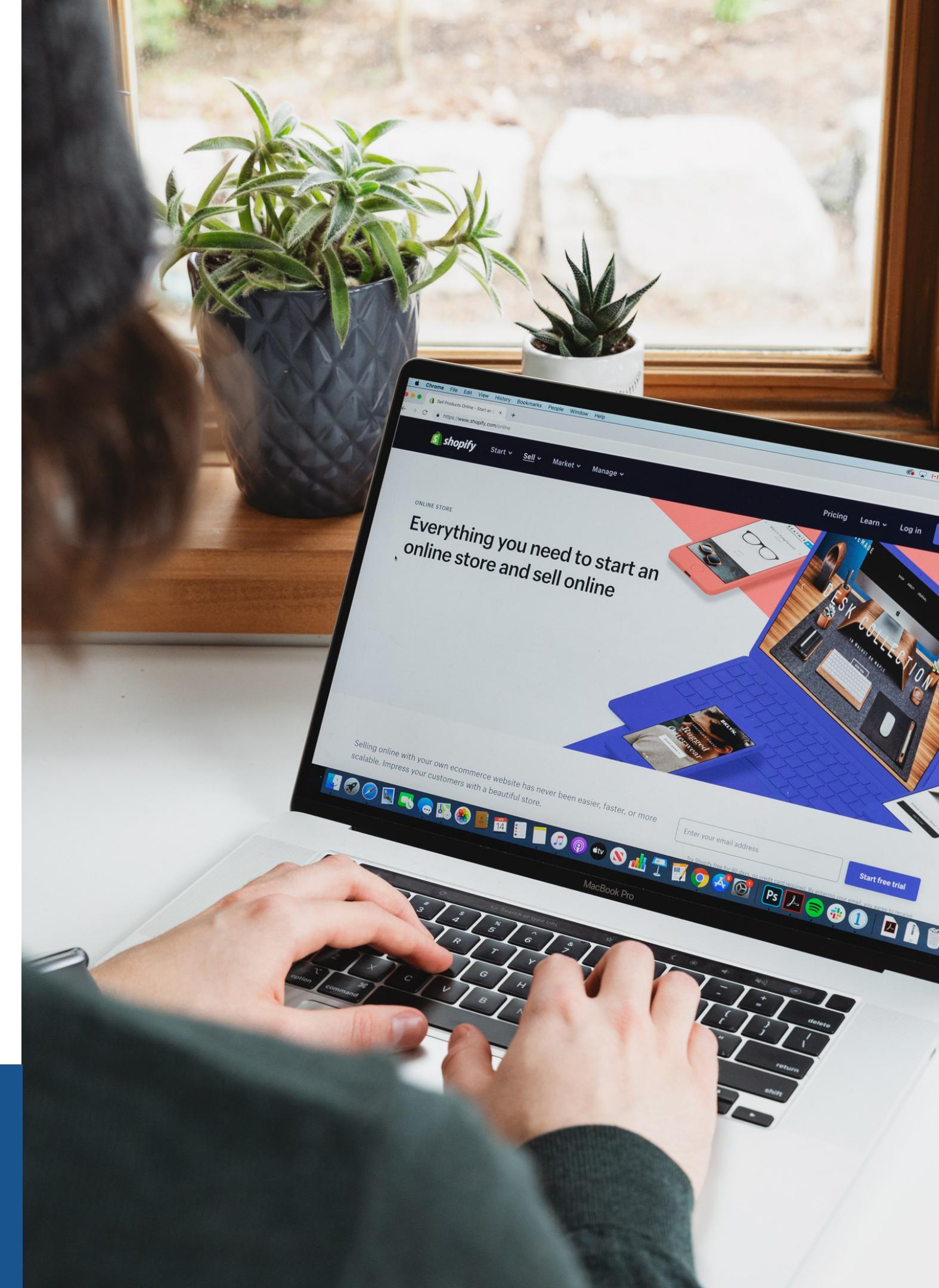


1

Pendahuluan

1.2 Tujuan

Tujuan dari dilakukannya Exploratory Data Analysis adalah untuk memahami performa dari e-commerce di Pakistan dan memberikan rekomendasi dasar untuk meningkatkan keuntungan berupa sales atau mengurangi biaya penjualan bagi e-commerce tersebut



1

Pendahuluan

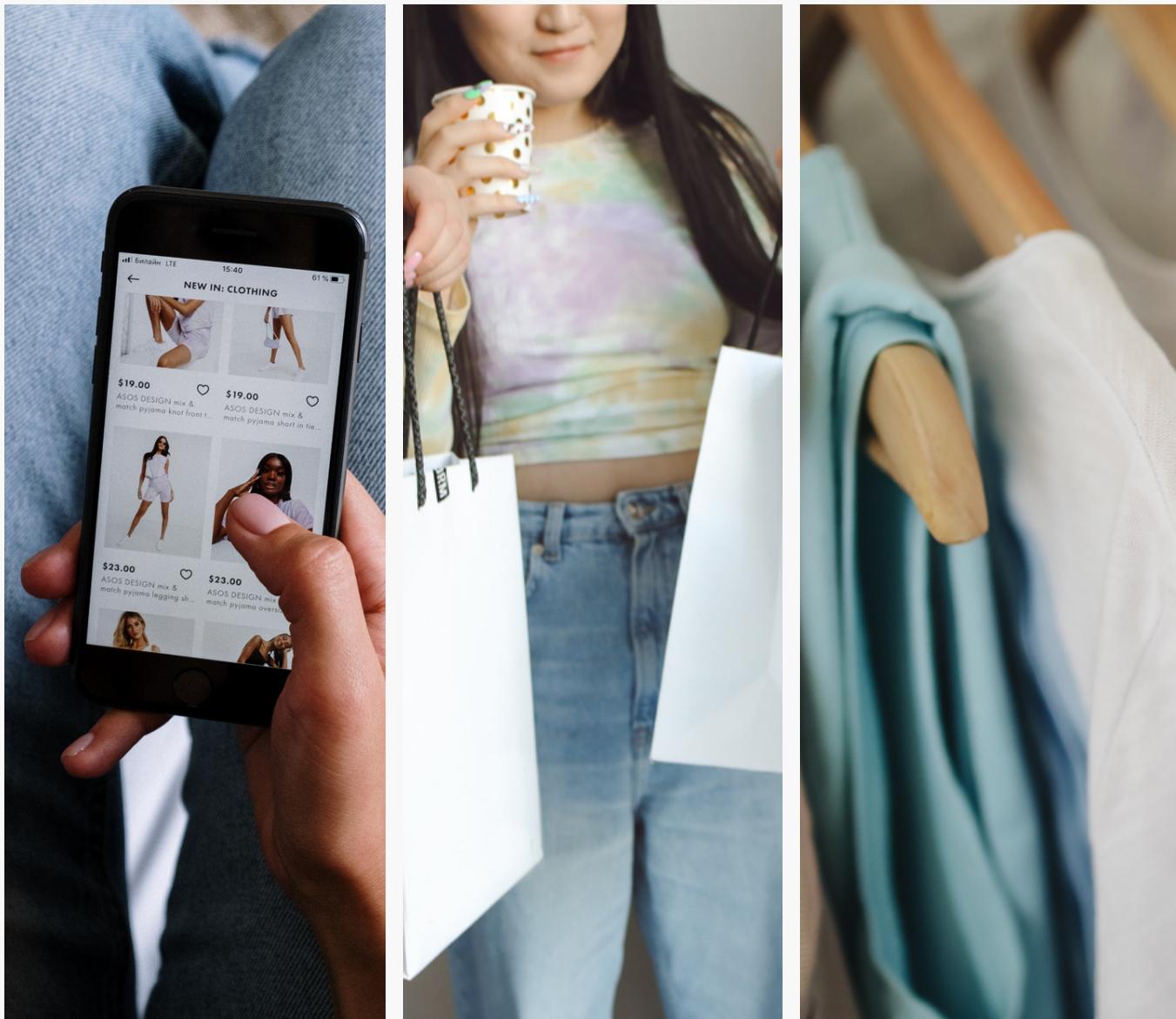
1.3 Batasan Masalah

1. Menganalisa kegiatan Marketing pada dataset E-commerce Pakistan
2. Menganalisa Produk dan Order pada dataset E-commerce Pakistan
3. Menganalisa Customers pada dataset E-commerce Pakistan



1

Pendahuluan



1.3 Deskripsi Variabel

- Item_id : Nomor unik di setiap transaksi
- Status : Status transaksi
- Created_at : Tanggal & waktu dilakukannya pemesanan barang
- Sku : Stock Keeping Unit dari setiap produk yang terjual
- Price : Harga tiap unit dari produk yang dijual
- Qty_ordered : Jumlah total unit barang yang dipesan
- Grand_total : Total harga keseluruhan yang dibayar oleh customer
- Increment_id : kode struk
- Category_name_1 : Kategori barang yang terjual
- Sales_commission_code : Kode komisi penjualan
- Discount_amount : jumlah diskon yang diberikan
- Payment_method : metode pembayaran
- Working_date : Tanggal & waktu setiap transaksi
- BI_Status : Status dari transaksi pada sistem Business Intelligence
- MV : Monetary Value dari setiap transaksi
- Year : Tahun customer bergabung
- Month : Bulan customer bergabung
- Customer_since : informasi tanggal,bulan dan tahun customer bergabung
- M-Y : informasi tentang bulan dan tahun terjadinya transaksi
- FY : tahun fiskal perusahaan
- Customer_ID : Nomor customer (Primary Key)

2

Struktur Data

	Features	DataType	Null	NullPercentage	Unique
0	item_id	float64	464051	44.26	584524
1	status	object	464066	44.26	16
2	created_at	object	464051	44.26	789
3	sku	object	464071	44.26	84889
4	price	float64	464051	44.26	9121
5	qty_ordered	float64	464051	44.26	74
6	grand_total	float64	464051	44.26	36829
7	increment_id	object	464051	44.26	408785
8	category_name_1	object	464215	44.27	16
9	sales_commission_code	object	601226	57.34	7226
10	discount_amount	float64	464051	44.26	28058
11	payment_method	object	464051	44.26	18
12	Working Date	object	464051	44.26	789
13	BI Status	object	464051	44.26	4
14	MV	object	464051	44.26	9720
15	Year	float64	464051	44.26	3
16	Month	float64	464051	44.26	12
17	Customer Since	object	464062	44.26	26
18	M-Y	object	464051	44.26	26
19	FY	object	464051	44.26	3
20	Customer ID	float64	464062	44.26	115326
21	Unnamed: 21	float64	1048575	100.00	0
22	Unnamed: 22	float64	1048575	100.00	0
23	Unnamed: 23	float64	1048575	100.00	0
24	Unnamed: 24	float64	1048575	100.00	0
25	Unnamed: 25	float64	1048575	100.00	0

Import Library

Library / package yang digunakan adalah

- Numpy
- Pandas
- Seaborn
- Matplotlib pyplot
- Operator (attrgetter)

Bentuk Data

- Data memiliki 26 kolom dan 1048575 baris
- Data memiliki value NaN sebesar 45%
- Customer ID merupakan Primary Key dari data
- Data numerikal berjenis Float
- Data kategorikal berjenis Object (campuran)

3

Data Preparation

Menghitung dan Menghapus data terduplicasi

Data duplikat tidak dibutuhkan dalam Exploratory Data Analysis, oleh karenanya data tersebut dihapus.

Menghapus kolom yang tidak dibutuhkan

Kolom dengan nama Unnamed tidak memiliki isi sama sekali dan memberikan persentase missing value sebesar 100%. Oleh karenanya dilakukan penghapusan

Menghitung dan menghapus data anomali & missing value

Missing value adalah tipe data yang tidak ada isinya. Data anomali adalah basis data yang memberikan efek samping yang tidak diharapkan oleh karenanya kita dapat mengisinya atau menghapus dari data tersebut

Fillna kolom sales commission code & category

- Pada kolom kategori banyak berisi value '\\N' sebesar 7850 data sedangkan value NaN sebesar 165. Setelah diselidiki, bahwa beberapa SKU dari setiap transaksi dapat merepresentasikan jenis kategori barang yang dibeli.
- Pada kolom sales_commission_code banyak berisi value '\\N' sebesar 339001 data sedangkan value NaN sebesar 137176. Kami berasumsi bahwa values tersebut mengindikasikan tidak adanya penggunaan kode komisi untuk sales. Sehingga kami isikan data tersebut dengan 'no commission'

3

Data Preparation

Membuat kolom Grand Total baru

Terdapat anomali pada value dari kolom grand total.

	item_id	status	created_at	sku	price	qty_ordered	grand_total
17	211150.0	complete	7/1/2016	UK_Gift Box Mix Dry Fruit Sweets 500 Gms	360.0	1.0	1270.0

Grand total sendiri merupakan total harga yang dibayarkan oleh customer. Grand total memiliki rumus:
 $\text{grand_total} = (\text{price} * \text{qty_ordered}) = \text{discount amount}$

Membuat kolom Revenue

Pada kolom grand total tidak mencerminkan apakah total harga tersebut merupakan revenue / pendapatan bagi perusahaan ecommerce tersebut dikarenakan kondisi status transaksi yang terjadi. Sehingga dibutuhkan kolom revenue berdasarkan transaksi yang sudah selesai / completed

3

Data Preparation

Membuat kolom potential revenue

Potential revenue merupakan perkiraan revenue yang dapat didistribusikan sebagai revenue di masa depan. Potensial revenue tidak meliputi status transaksi yang sudah lengkap / completed.

Membuat kolom loss_revenue

Loss revenue merupakan pendapatan / revenue yang gagal didapatkan oleh perusahaan, dalam hal ini e-commerce pakistan. Loss revenue dalam hal ini dikategorikan berdasarkan status / kegiatan transaksi yang gagal. Diantaranya : canceled, fraud, closed

Membuat kolom Date Feature

Penggunaan date feature memudahkan untuk menganalisis topik yang memiliki dimensi waktu. Diantaranya cohort analysis, RFM analysis, dan lain lain

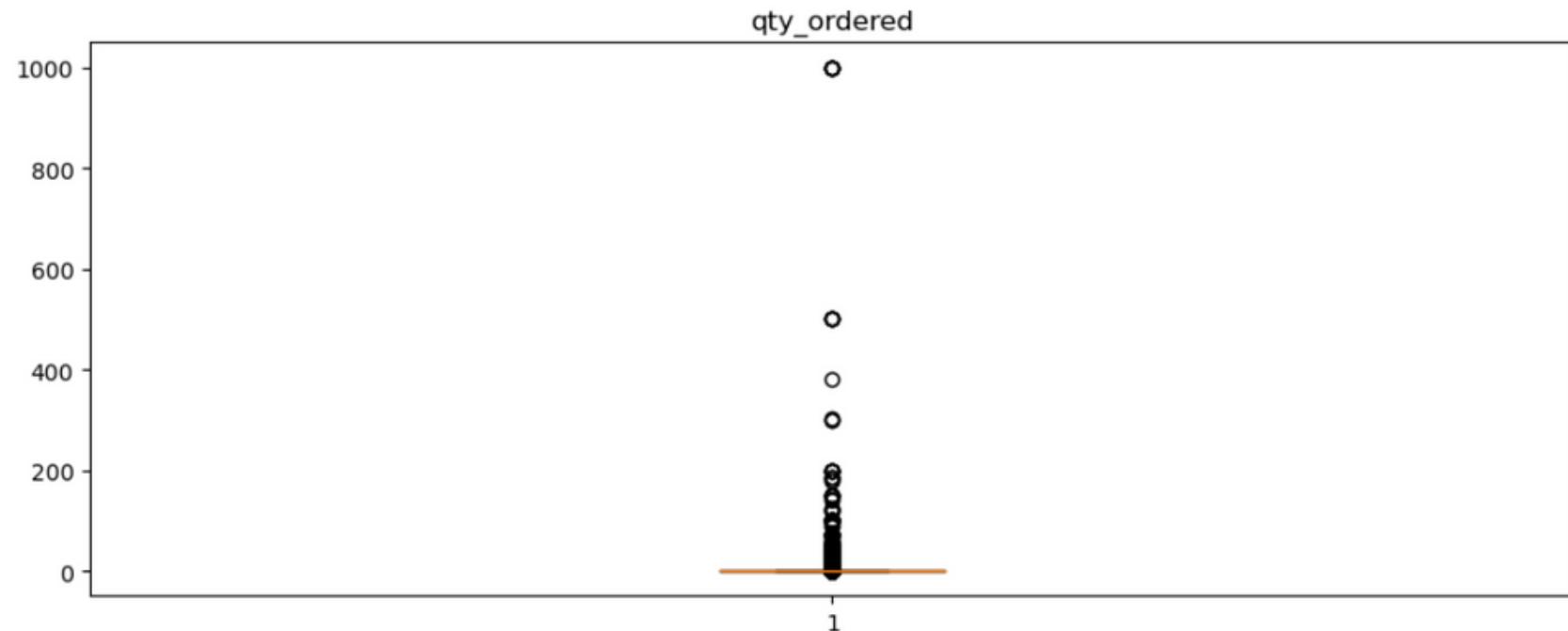
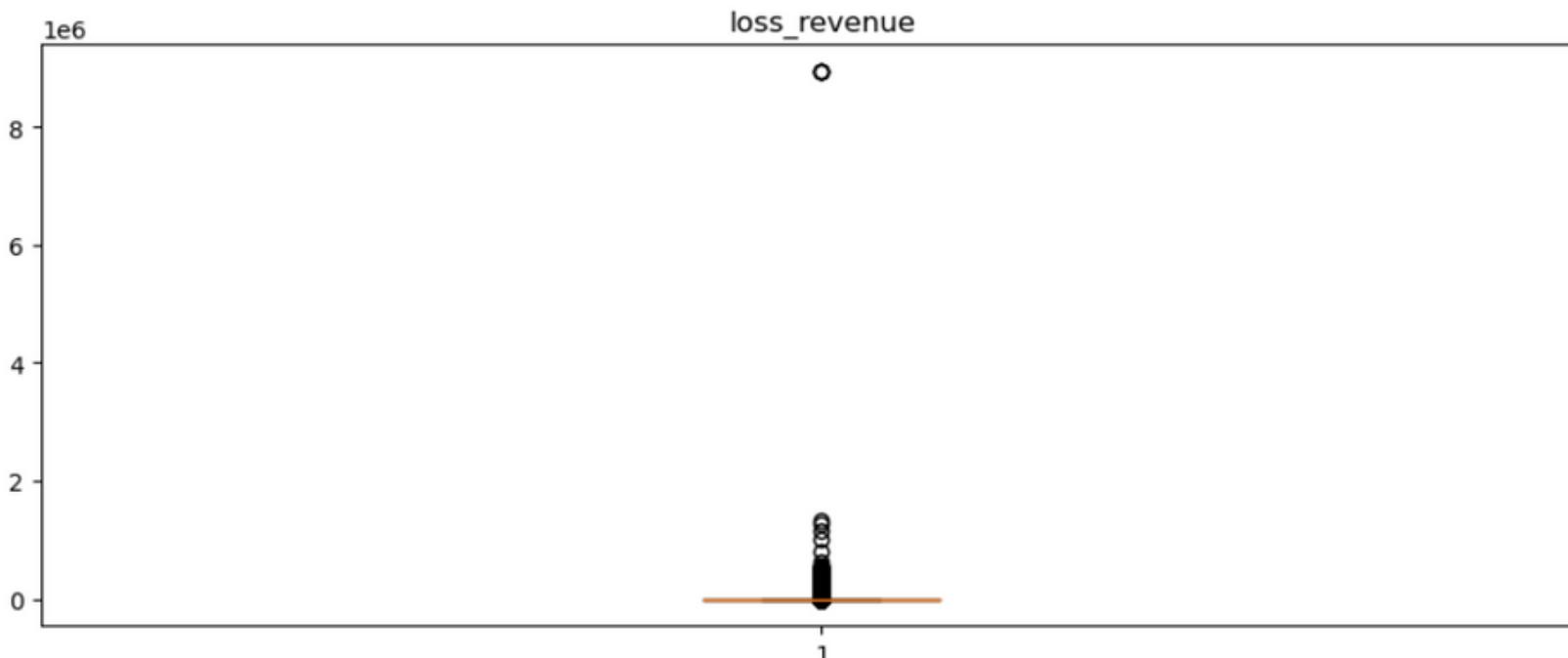
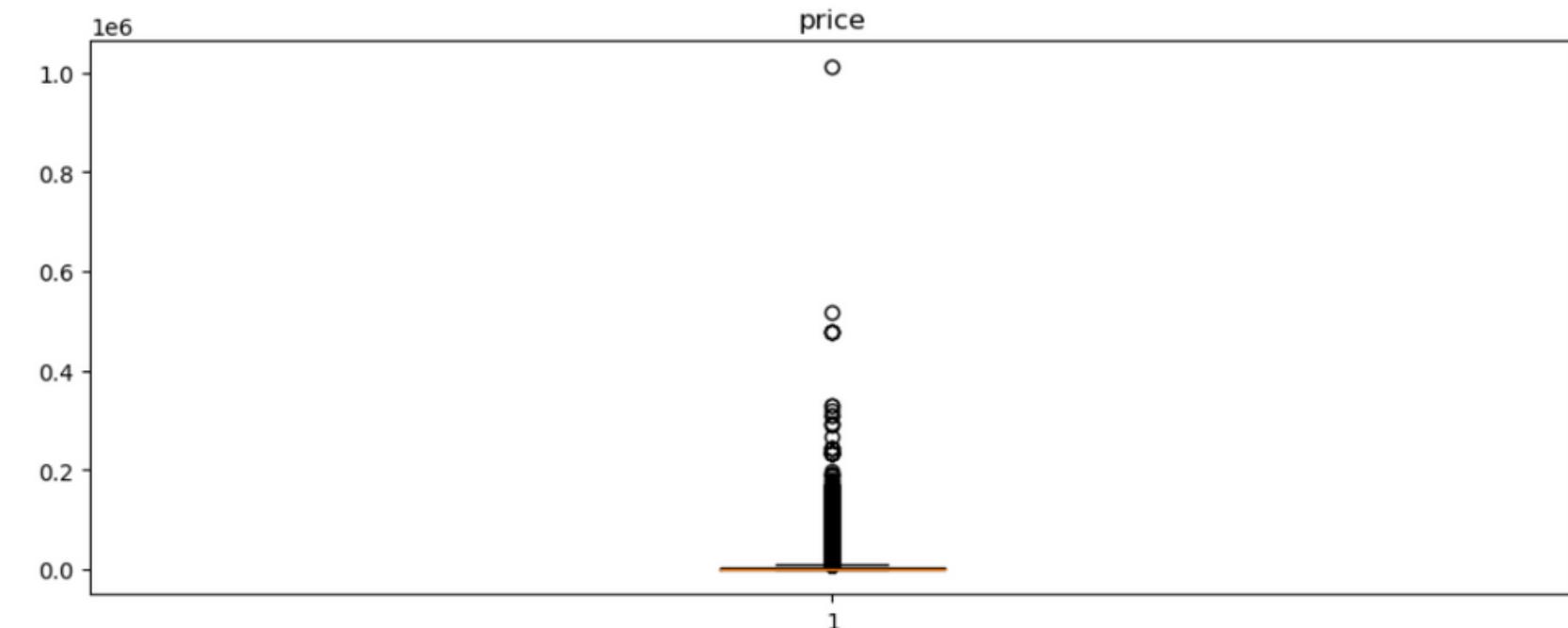
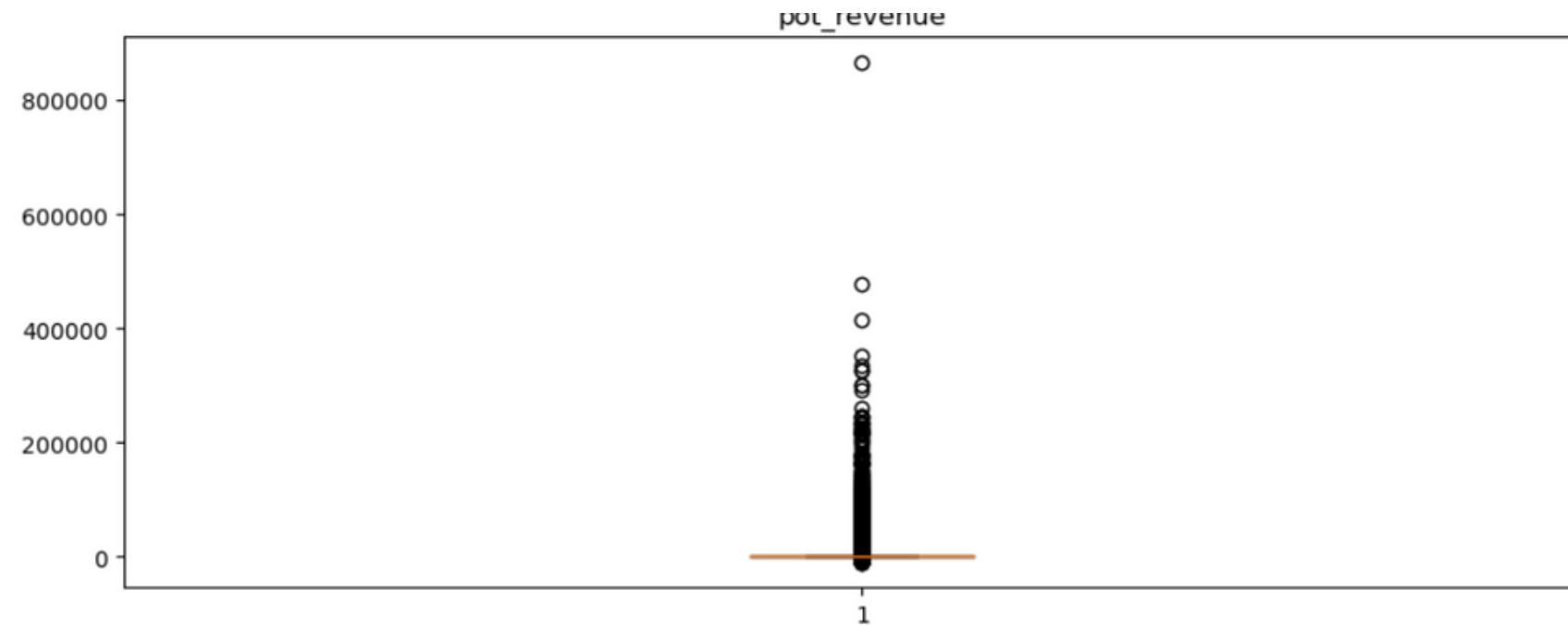
4.1 Analisis Deskriptif Data Numerikal

	count	mean	std	min	25%	50%	75%	max
item_id	576764.0	568941.105932	199131.165609	211133.00	400906.75	572793.5	741249.500	905208.0
price	576764.0	6377.588553	15003.885937	0.00	360.00	900.0	4053.000	1012625.9
qty_ordered	576764.0	1.297281	4.005476	1.00	1.00	1.0	1.000	1000.0
discount_amount	576764.0	503.075657	1510.340571	-599.50	0.00	0.0	164.850	90300.0
year	576764.0	2017.054769	0.704204	2016.00	2017.00	2017.0	2018.000	2018.0
month	576764.0	7.149080	3.488661	1.00	4.00	7.0	11.000	12.0
customer_id	576764.0	46233.241184	34394.731631	2.00	13929.00	43603.5	74036.000	115326.0
grand_total	576764.0	6711.611218	33856.505058	-15410.00	350.00	949.0	4500.000	8944000.0
revenue	230457.0	4086.432185	28146.582130	-15410.00	300.00	722.5	2200.000	8944000.0
pot_revenue	576764.0	640.880307	5684.296906	-9888.65	0.00	0.0	0.000	866416.0
loss_revenue	576764.0	3991.208063	28269.950215	-10087.00	0.00	0.0	620.425	8944000.0
dateofmonth	576764.0	17.563097	8.078875	1.00	11.00	19.0	24.000	31.0
week	576764.0	29.431766	15.278425	1.00	16.00	30.0	46.000	52.0

4.2 Analisis Deskriptif Data Kategorikal

		count	unique	top	freq
	status	576764	15	complete	230457
	sku	576764	84611	MATSAM59DB75ADB2F80	3775
	increment_id	576764	402113	100266667	72
	category	576764	15	Mobiles & Tablets	115710
	sales_commission_code	576764	7175	no commission	470023
	payment_method	576764	18	cod	267981
	bi_status	576764	3	Net	230949
	mv	576764	9701	999	9498
	customer_since	576764	26	2016-11	80891
	m-y	576764	26	11-2017	83928
	576764	3	FY18	306498	
	ctg	576764	1259	MAT	61202
	quarter	576764	9	2017Q4	117173
	Date	576764	789	2016-11-25	14929

4.3 Pengukuran sebaran dan outlier





Pembahasan

01

Analisa Marketing

- Analisa Revenue
- Analisa Average Order
- Analisa Revenue per Customer
- Analisa Number of Orders
- Analisa Number of Customer
- Top 10 commission code

02

Analisa Produk dan Order

- Revenue vs Potensial
- Revenue vs Loss Revenue
- Orders
- Top Payment
- Korelasi

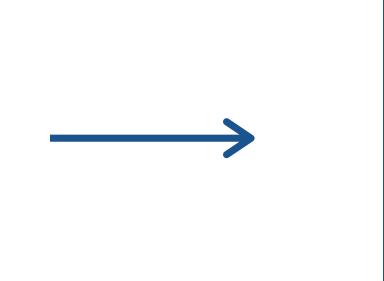
03

Analisa Customer

Analisis RFM (Recency, Frequency, Monetary) untuk segmentasi customer Ecommerce



Analisis Marketing



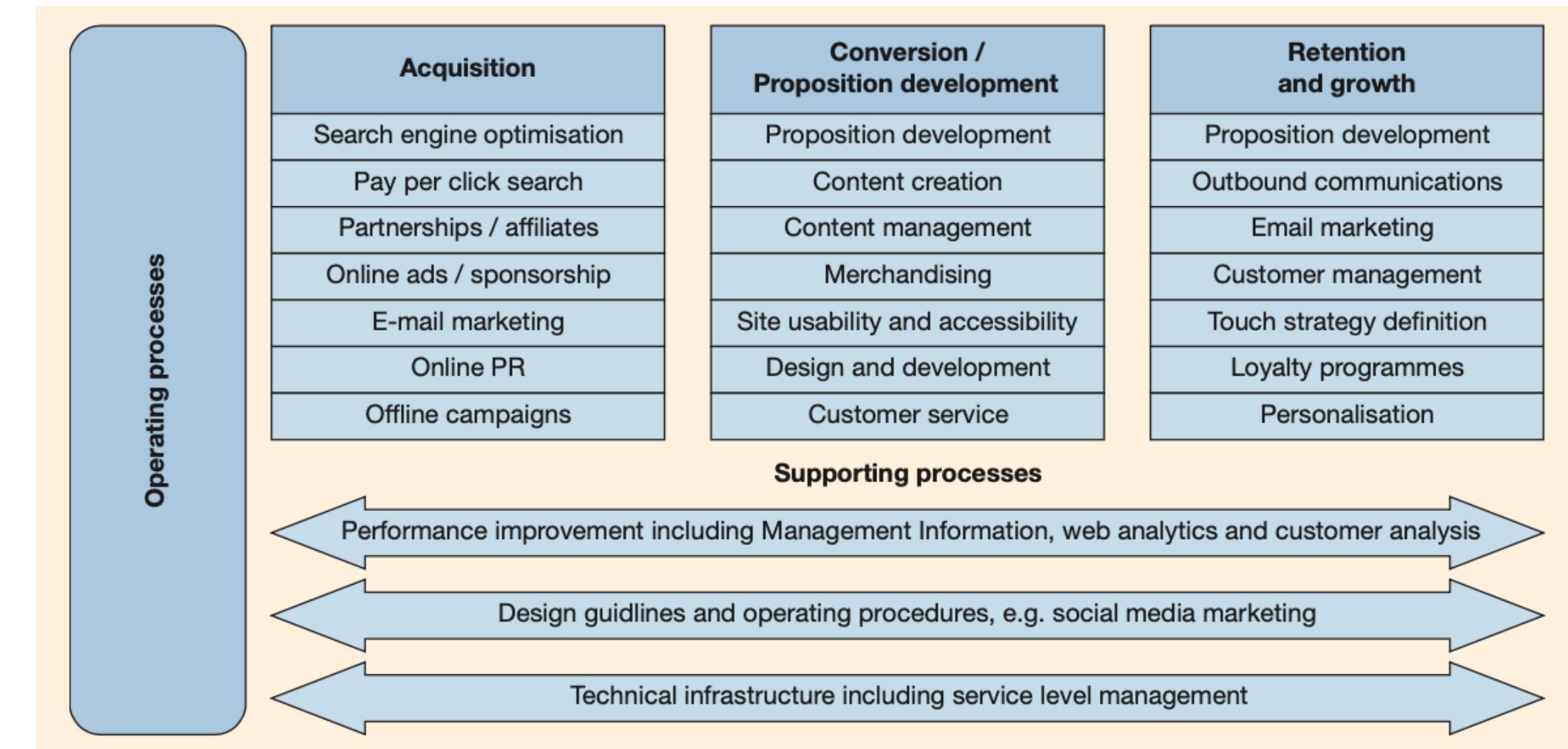
MARKETING



3.1 Analisa Marketing

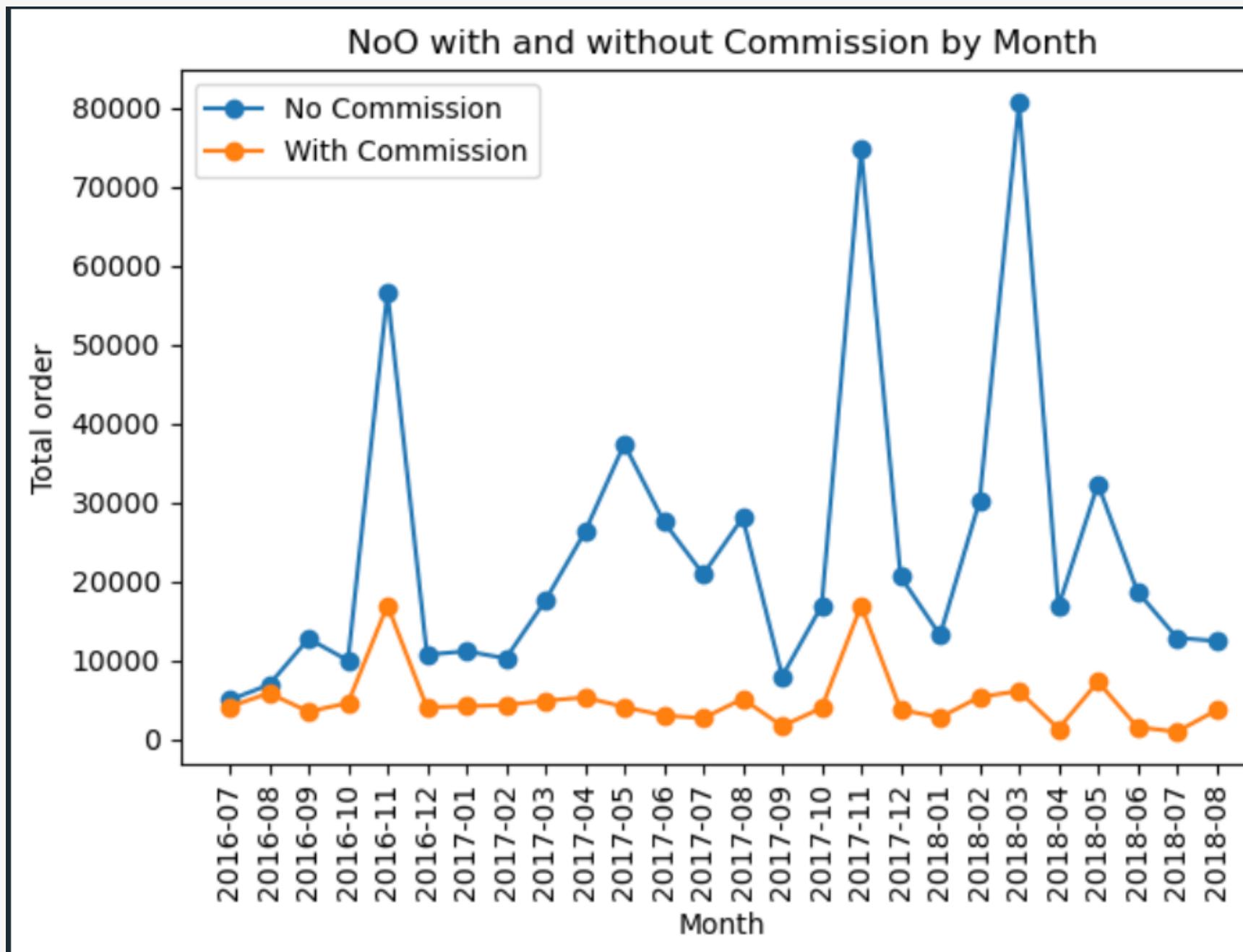
Digital marketing adalah suatu strategi pemasaran menggunakan media digital dan internet. Berikut adalah beberapa tahapan umum mengenai marketing digital:

- Customer Acquisition
- Customer Conversion
- Customer Retention & Growth



Adanya indikasi penggunaan sales commission code merupakan kegiatan pada tahapan Customer Conversion / Customer Retention & Growth

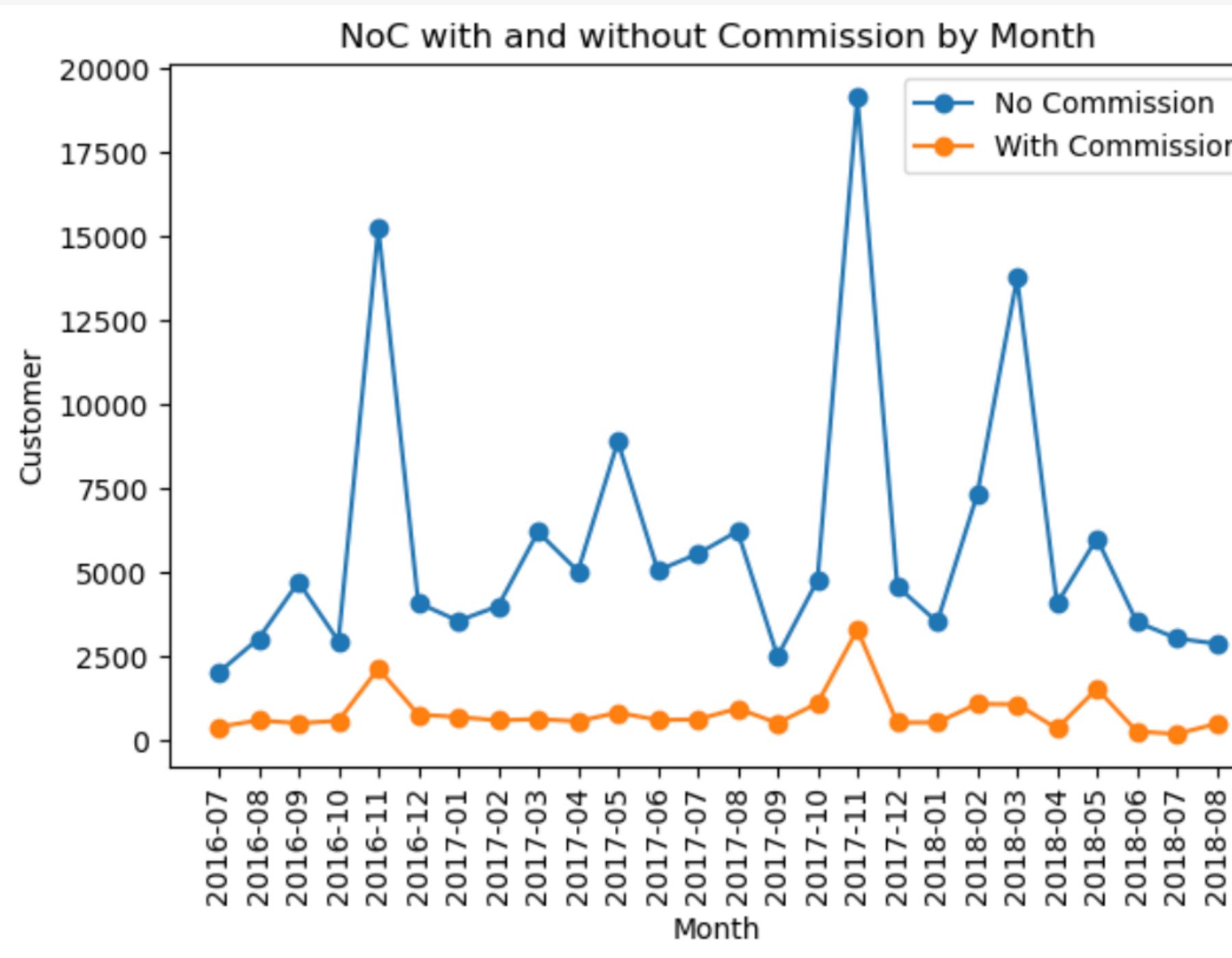
Number of Order berdasarkan penggunaan kode komisi



ANALYSIS:

- Total order tanpa kode komisi lebih tinggi dibandingkan penggunaan kode komisi
- Total order tanpa komisi lebih berfluktuatif
- Penggunaan kode komisi relatif lebih stabil
- Jumlah order tertinggi tanpa kode komisi terdapat pada bulan :
 - 03-2018
 - 11-2017
 - 11-2016

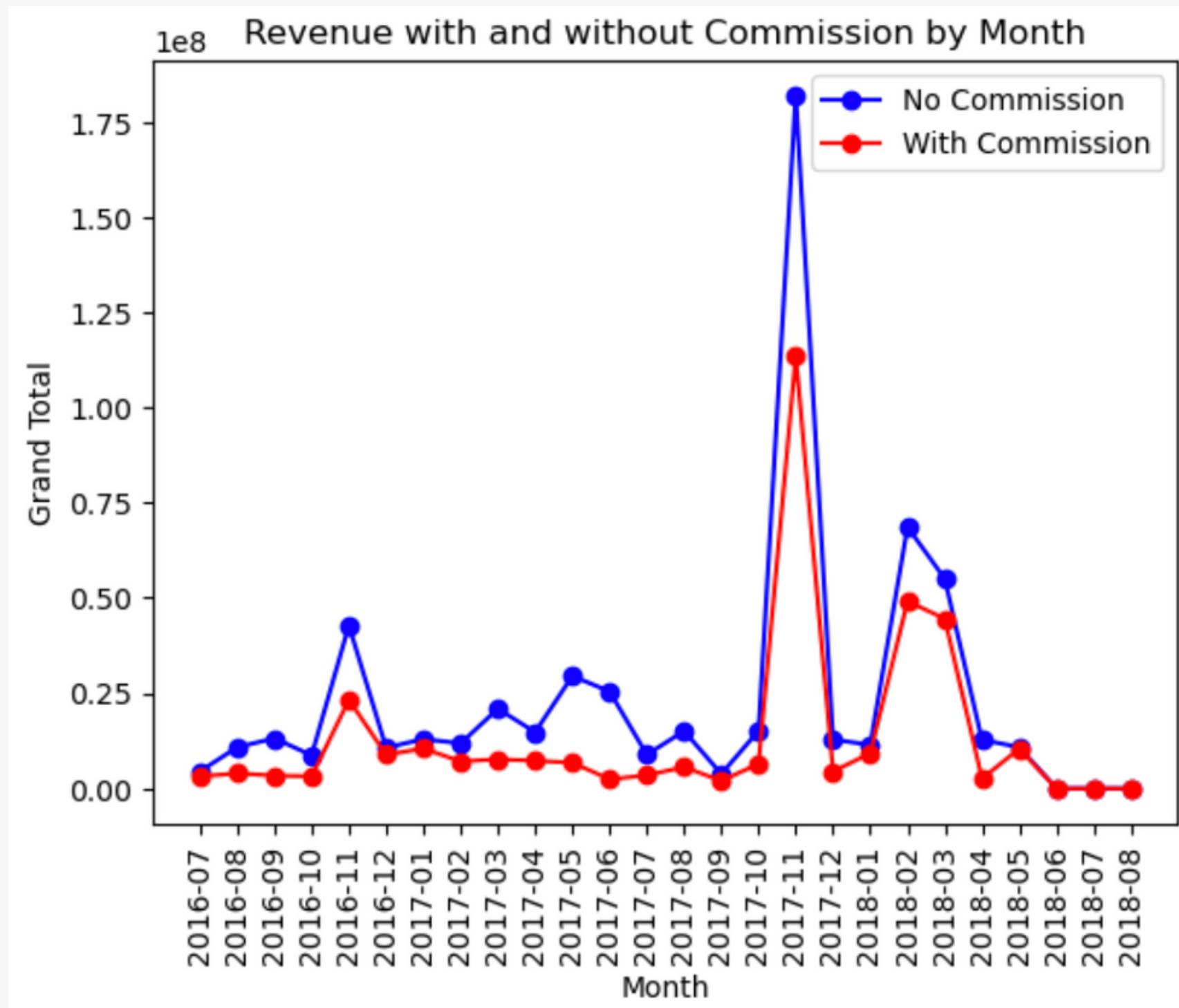
Number of Customer berdasarkan penggunaan kode komisi



ANALYSIS:

- Total customer tanpa kode komisi lebih tinggi dibandingkan penggunaan kode komisi
- Total customer tanpa komisi lebih berfluktuatif
- Customer dengan penggunaan kode komisi relatif lebih stabil
- Jumlah customer tertinggi tanpa kode komisi terdapat pada bulan :
 - 11-2017
 - 03-2018
 - 11-2016

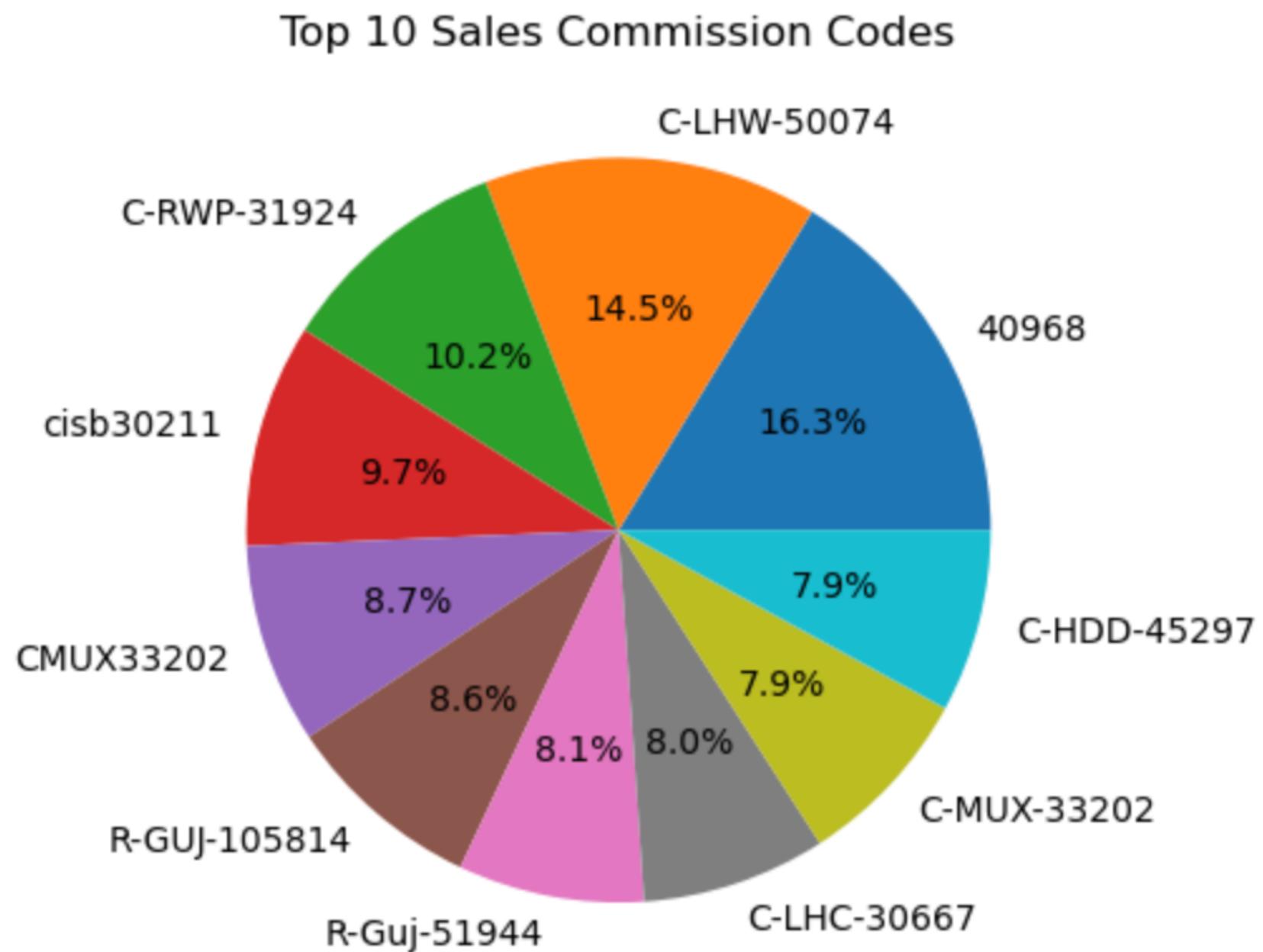
Revenue berdasarkan penggunaan kode komisi



ANALYSIS:

- Revenue tanpa kode komisi lebih tinggi dibandingkan penggunaan kode komisi. Hal ini sesuai dengan tingginya grafik order dan customer
- Grafik antara kedua komisi tersebut relatif sama. Dibutuhkan strategi lebih agar revenue tanpa kode komisi lebih tinggi dibandingkan dengan kode komisi

Top 10 Sales Commission Code



ANALYSIS:

- Pie chart tersebut merupakan 10 sales commission code yang paling sering dipakai oleh customer
- Commission code tersebut dapat dijadikan referensi dan dianalisis untuk ditingkatkan

Marketing Insights & Rekomendasi

Strategi meningkatkan revenue:

- Cross selling : "Bagaimana dengan kombinasi pembelian handphone dengan casing handphone?"
- Upselling : "Apakah anda tertarik dengan menambahkan 10\$ untuk mendapatkan produk dengan kualitas yang lebih baik?"
- Free shipping : untuk produk yang memiliki harga tinggi
- Return Policy : "Garansi pengembalian jika barang tidak sesuai"

Melakukan AB Testing (cross selling):

A : shopping cart dengan rekomendasi "Orang yang membeli barang Y juga membeli barang Z"

B : shopping cart dengan rekomendasi "Berikut adalah rekomendasi produk untukmu"

Rekomendasi penambahan data :

Analaisis marketing ini hanya berdasarkan penggunaan kode komisi. Diperlukan tambahan informasi mengenai :

- Sumber traffic customer yang beraktifitas ke e-commerce (campaign analysis, Search Engine Marketing, Search Engine Optimizer, Affiliate Analytics, Email Marketing, dll)
- Cost to Acquire suatu consumer
- Gross Revenue per Customer (CLTV)
- Margin per Customer (untuk CLTV)

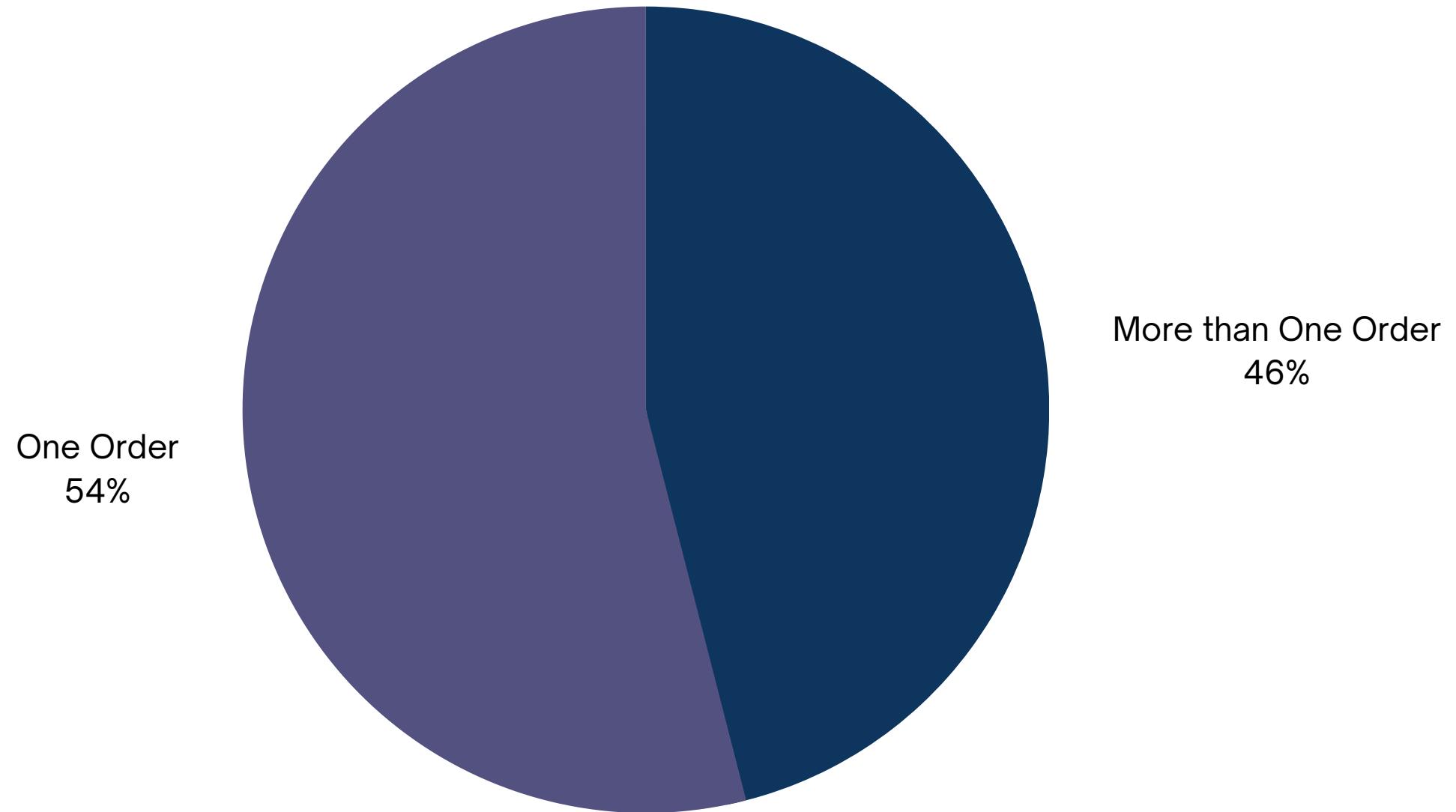


Analisis Produk dan Order



Order per Customer

Breakdown

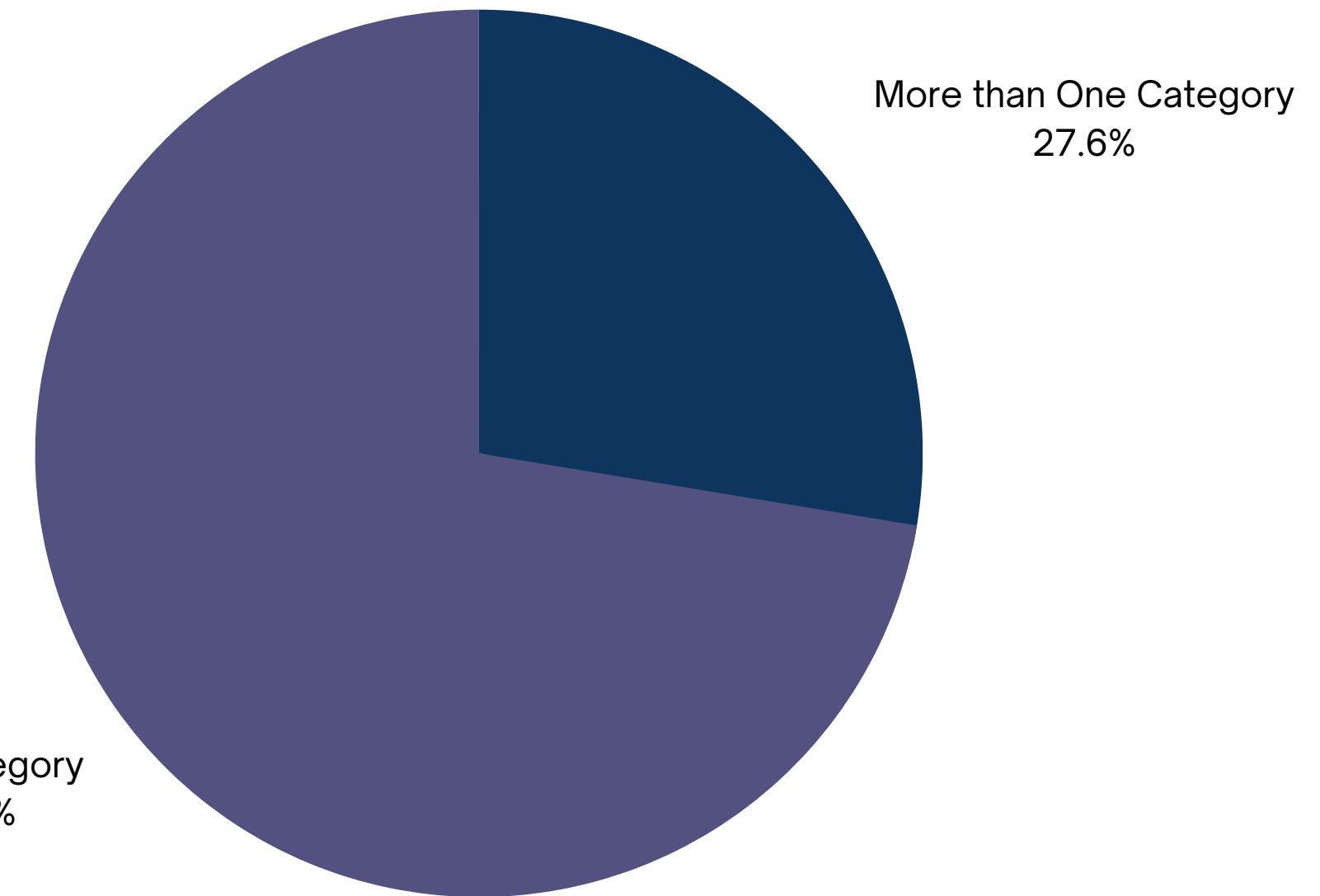


ANALYSIS & RECOMMENDATION

- Komposisi customer yang bertransaksi dengan 1 order dan lebih dari 1 order hampir sama
- Lakukan kegiatan "Cross Selling" untuk meningkatkan revenue

Orders per Category

Breakdown



ANALYSIS & RECOMENDATION

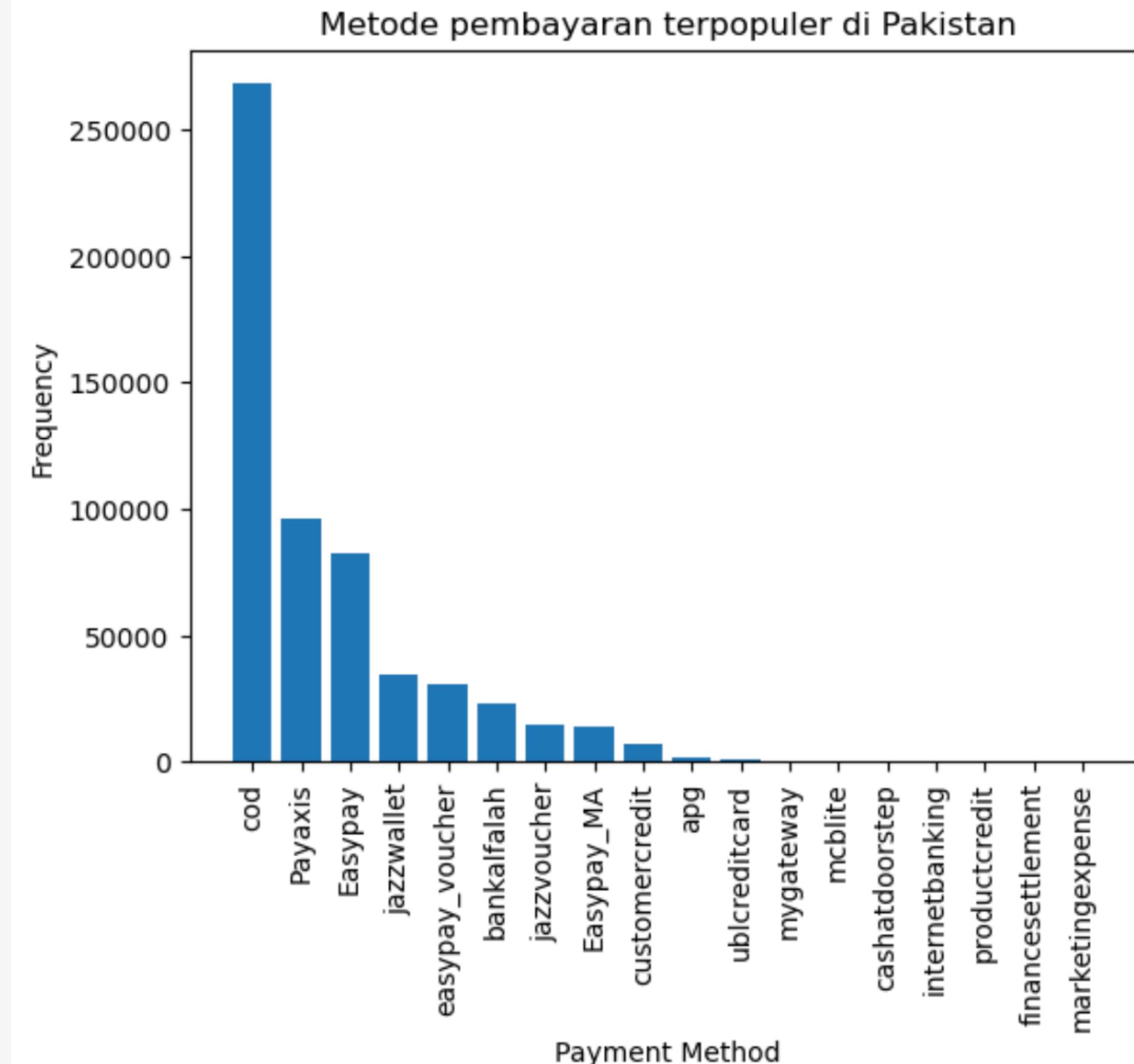
- Sekitar 2/3 customer yang bertransaksi di Ecommerce hanya berbelanja satu kategori
- Lakukan kegiatan "Cross Selling" atau Upselling

Metode Pembayaran terpopuler

ANALYSIS:

Cash on Delivery (COD) menjadi sumber metode pembayaran terbanyak dengan jumlah lebih dari 250.000 transaksi. Hal ini sejajar dengan informasi bahwa Pakistan merupakan cash-driven economy dikarenakan jumlah pemegang kartu debit/kredit dan akun m-wallet yang sangat rendah (Imtiaz et.al 2020)

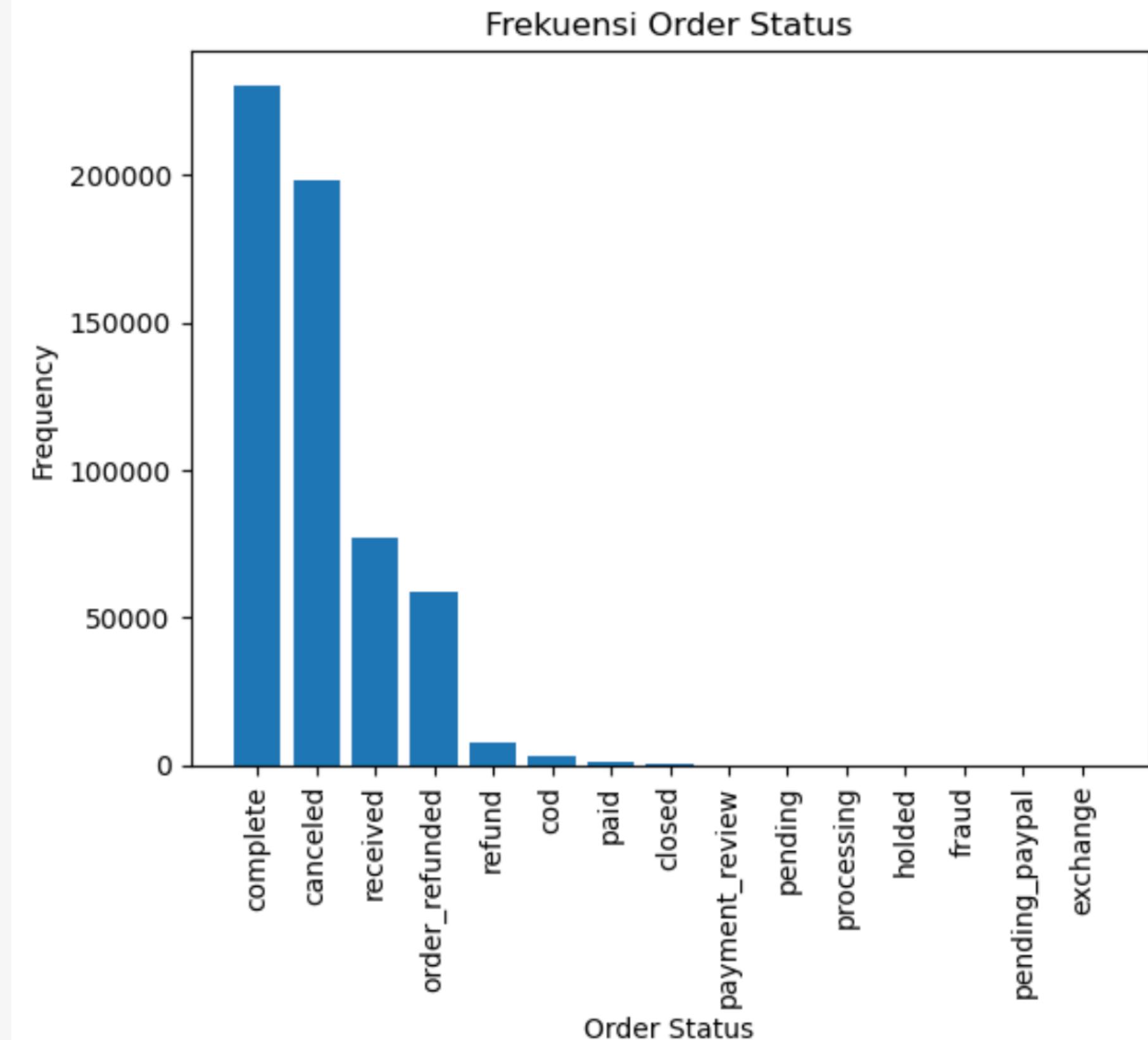
Tetapi hal perkembangan transaksi cash-less atau secara mobile sudah meningkat. Hal ini digambarkan dengan metode pembayaran yang terlihat di grafik



Frekuensi Status Order

ANALYSIS:

- Jumlah total transaksi yang selesai atau 'complete' lebih banyak dibandingkan yang lain
- Transaksi Canceled menjadi kategori terbanyak setelah complete. Perlu analisa lebih lanjut



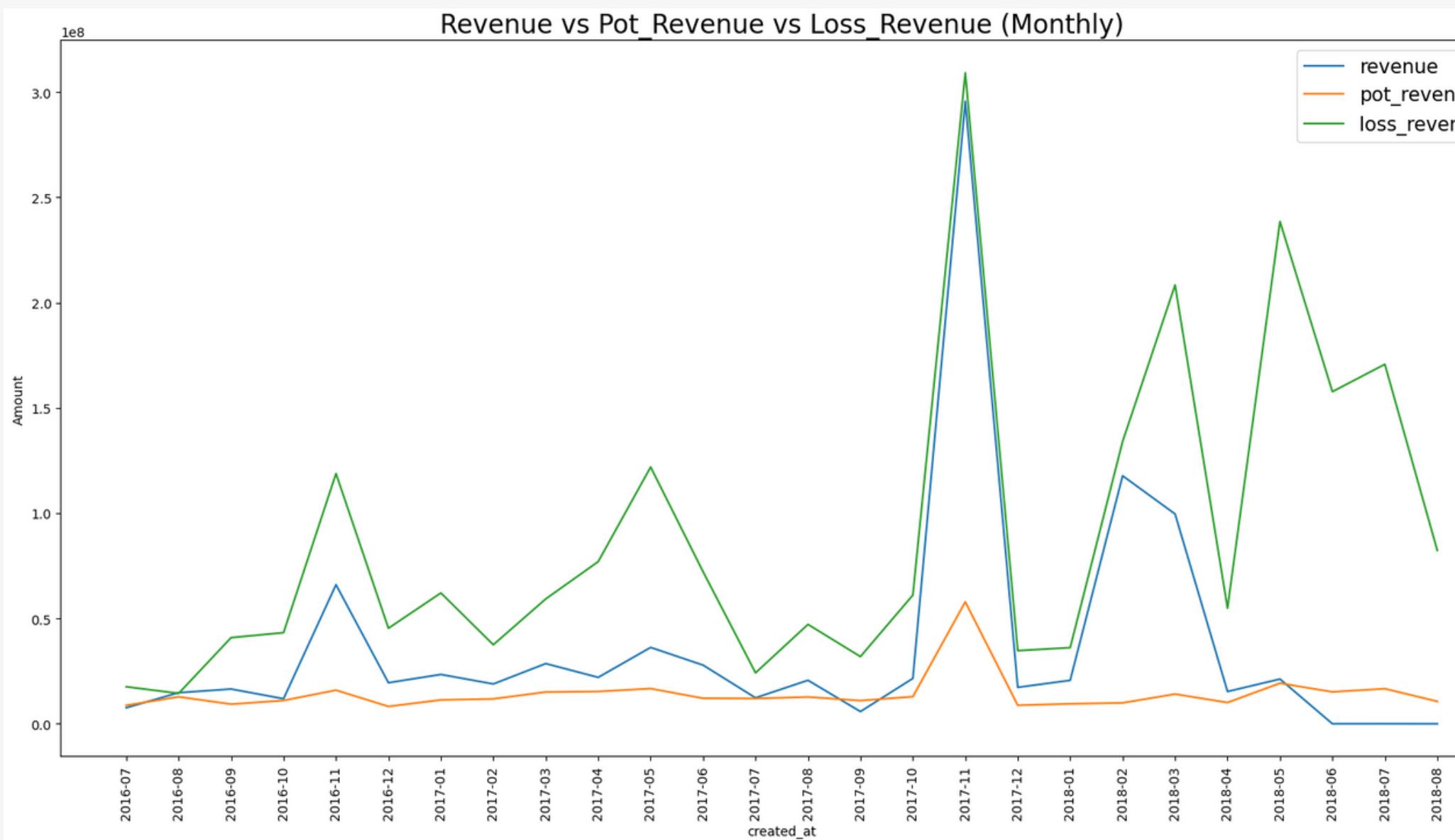
Revenue vs Potensial Revenue vs Loss Revenue

ANALYSIS

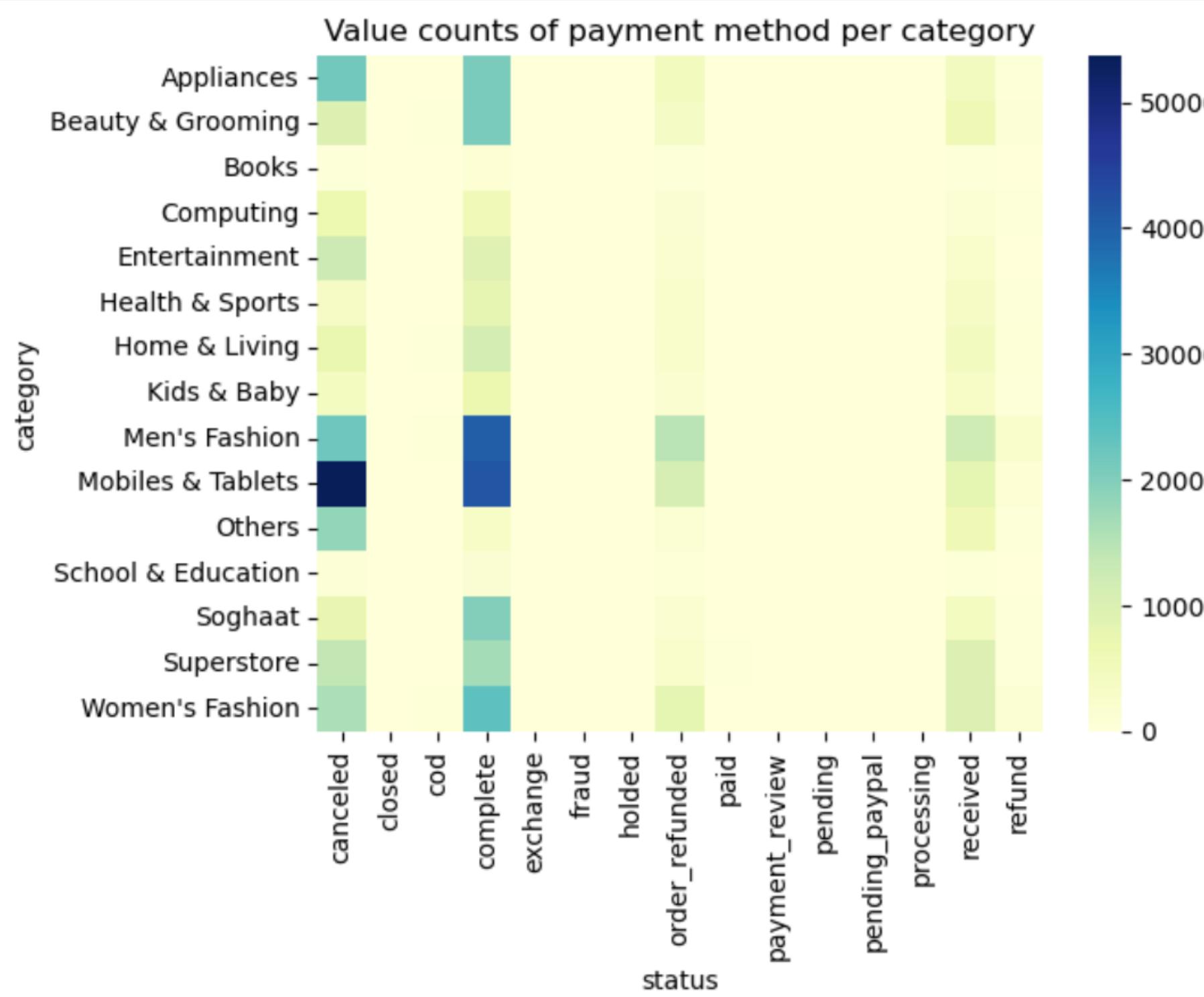
- Walaupun total status order "complete" lebih besar dibandingkan yang lain, tetapi Loss Revenue yang disebabkan oleh "canceled" order yang tinggi
- Loss Revenue merupakan angka yang bersifat "gambaran" bahwa perusahaan kehilangan kesempatan untuk mendapatkan profit. Bukan loss profit aktual

PERTANYAAN :

Apakah penyebab dari loss revenue?



Jumlah frekuensi status order berdasarkan Kategori



ANALYSIS UMUM:

Status Order "Completed" banyak terjadi pada kategori :

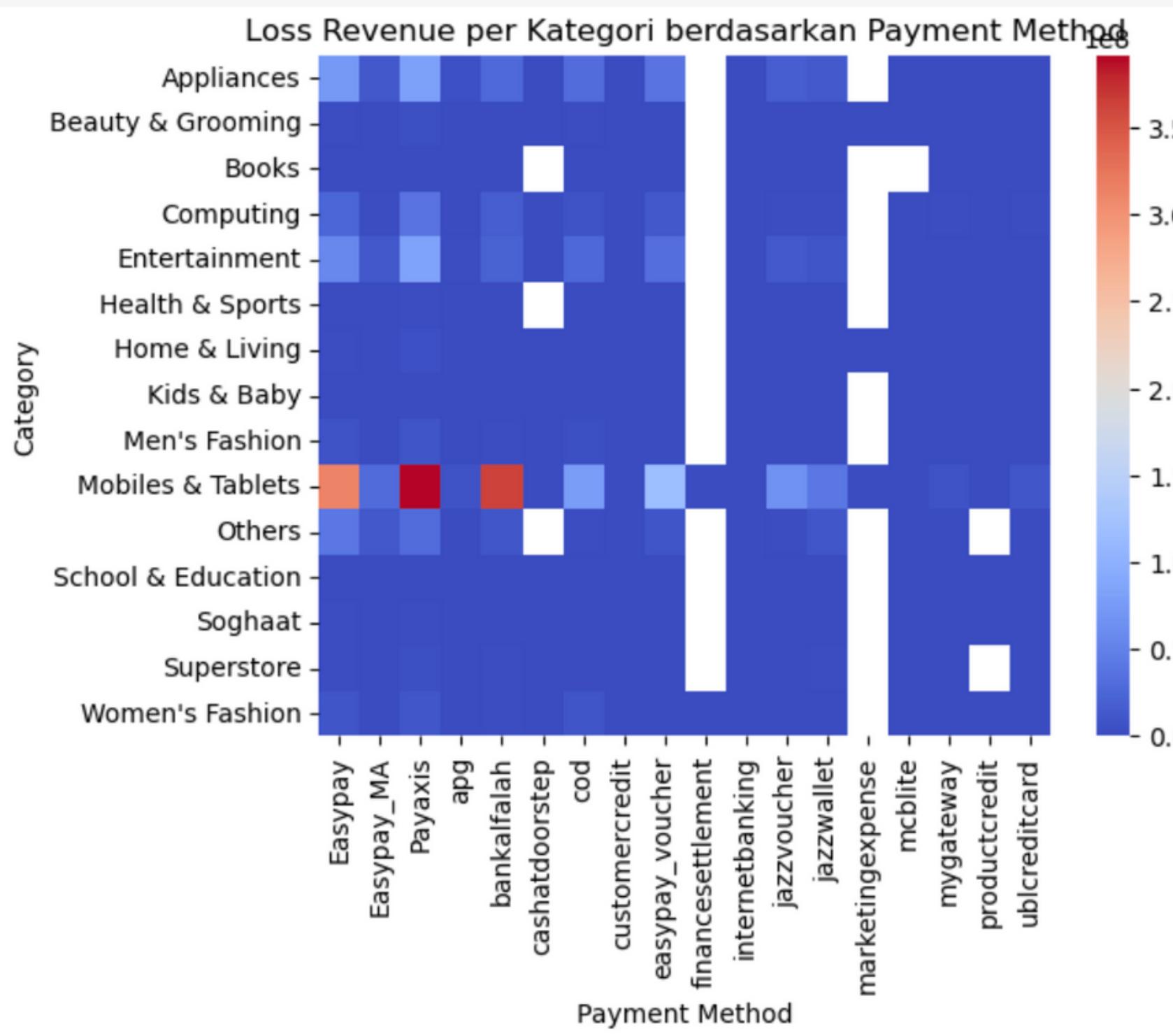
- Mobiles & Tablets (1)
- Men's Fashion (2)

ANALYSIS SPESIFIK:

Status order "Canceled" banyak terjadi pada kategori :

- Mobile & Tablets (1)
- Men's Fashion (2)
- Appliances (3) Order (3)

Loss Revenue per Kategori berdasarkan Payment Method



ANALYSIS UMUM:

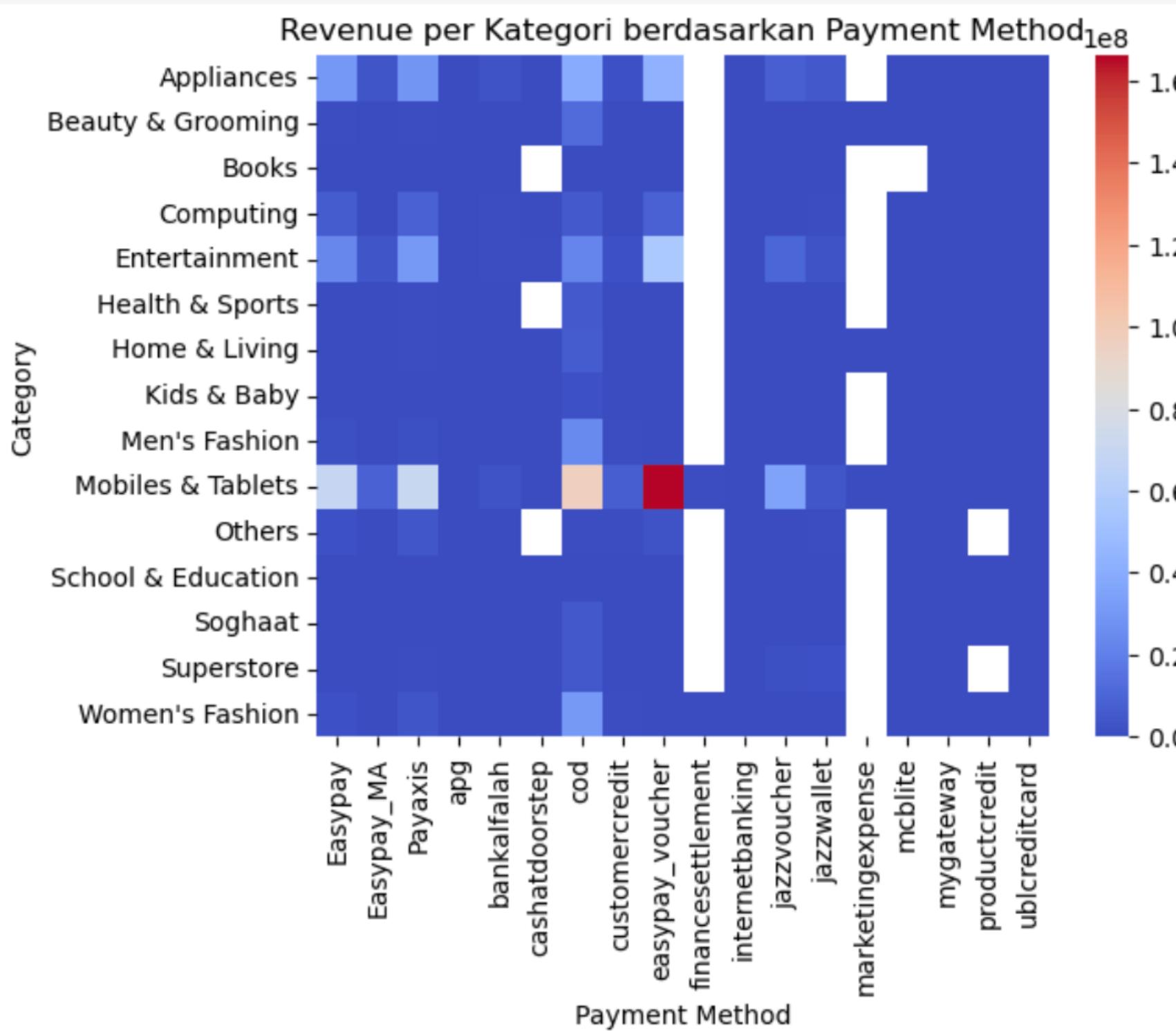
Loss Revenue yang memiliki karakteristik status order "canceled" dan "fraud" banyak terjadi pada kategori Mobile & Tablets, dengan metode pembayaran :

- payaxis (1)
- bankalfalah (2)
- easypay(3)

ANALYSIS SPESIFIK:

Salah satu penyebab terjadi banyaknya loss revenue disebabkan oleh metode pembayaran tersebut

Revenue per Kategori berdasarkan Payment Method



ANALYSIS UMUM:

Revenue yang memiliki karakteristik status order "completed" banyak terjadi pada kategori Mobile & Tablets, dengan metode pembayaran :

- easypay_voucher (1)
- COD (2)

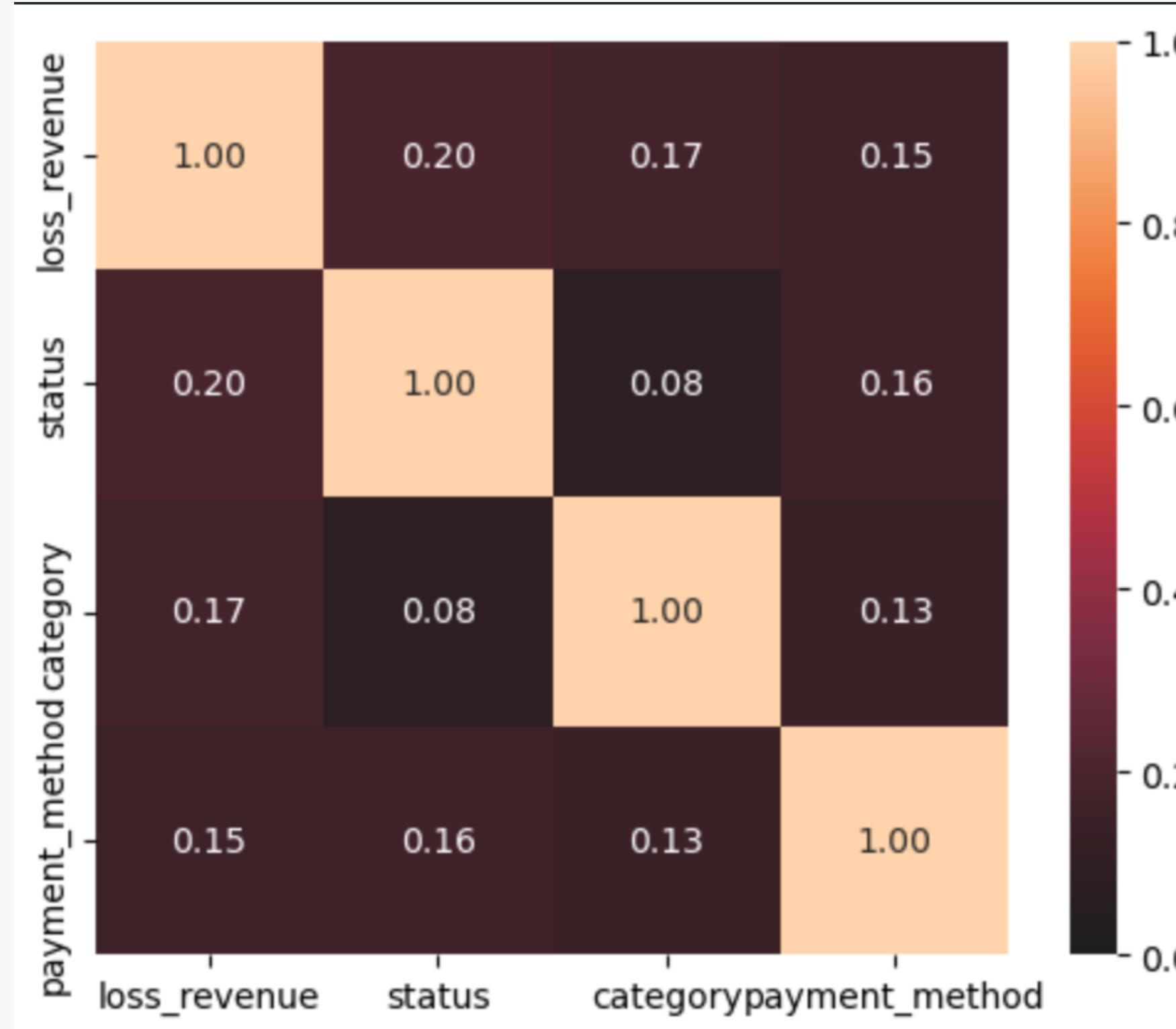
REKOMENDASI PENCEGAHAN LOSS REVENUE :

Loss Revenue mungkin dapat dimitigasi dengan menggunakan petode pembayaran

- Easypay_voucher
- COD

Dython Association Test

ANALYSIS UMUM:



- Hasil dari heatmap disamping menggunakan 4 variabel kategori dari dataset. Package / library yang digunakan adalah Cramer's V atau Cramer's phi dengan landasan perhitungan yaitu asosiasi antara variabel kategorikal
- Dari heatmap disamping menunjukan bahwa rendahnya tingkatan asosiasi antar variabel. Di mana range 0 menunjukan tidak adanya asosiasi sampai 1 yang menunjukan adanya asosiasi



Insight & Rekomendasi Produk dan Order

01

Step Completion

konversi dalam langkah yang merupakan titik transisi dalam pengalaman ecommerce yang merupakan bagian dari alur konversi

1. Landing page
2. Pencarian produk
3. Halaman produk
4. Order produk

02

Abandonment Analysis

analisis perilaku yang menyebabkan orang meninggalkan situs :

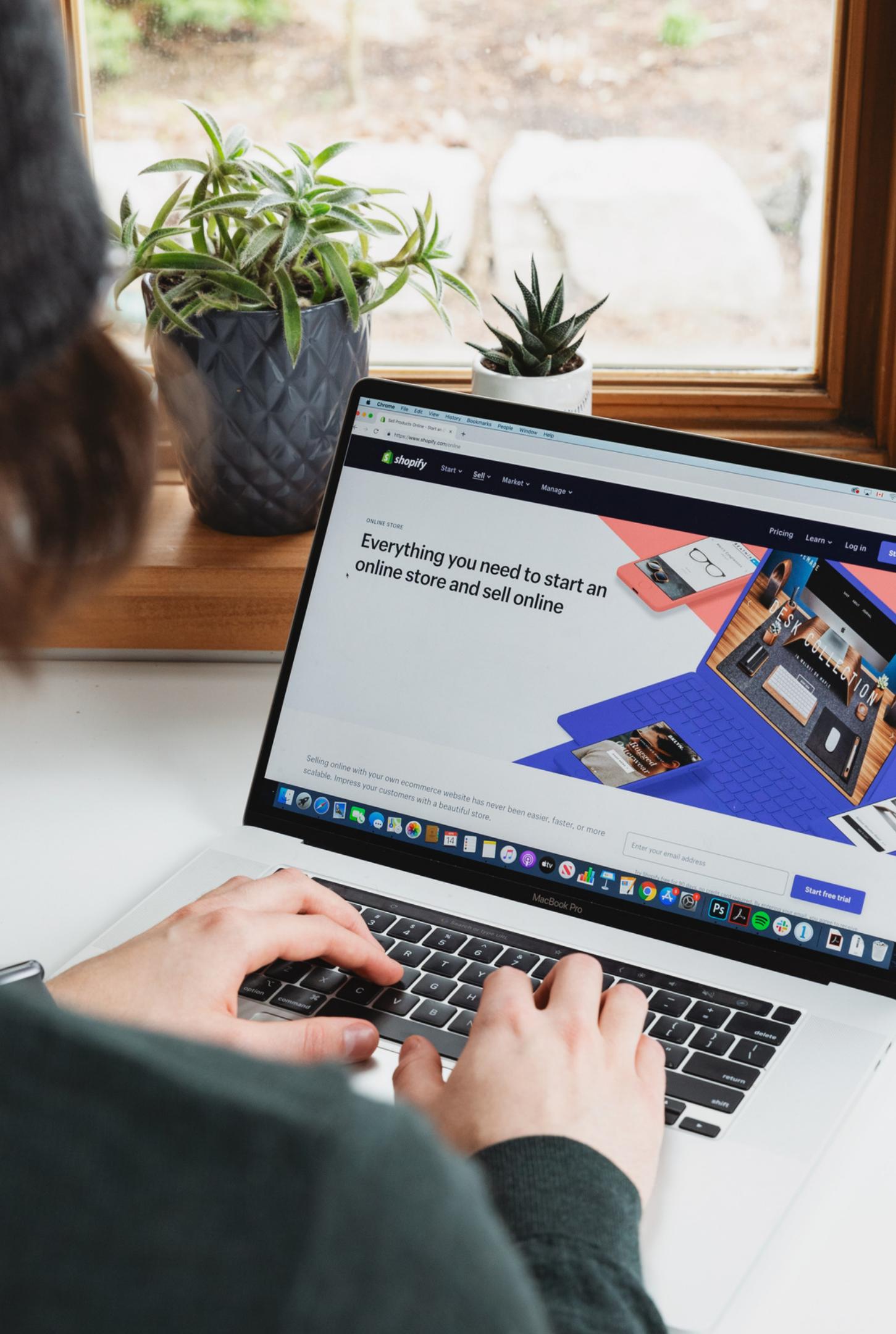
- Bounce rate
- Exit rate

03

In Page Analysis

Tautan, klik, sentuhan, dan kemungkinan interaksi lainnya dengan konten di halaman atau layar pengalaman e-niaga dapat diukur, dihitung, dan divisualisasikan

- Peta klik dan overlay halaman menunjukkan jumlah klik dan metrik terkait lainnya pada klik tertentu



RFM Segmentasi



RFM

APA ITU RFM ANALISIS?

Analisis ini biasanya dijadikan dasar untuk menyusun strategi pemasaran suatu perusahaan. Penerapan RFM analysis akan membantu perusahaan dalam melakukan segmentasi pelanggan secara efektif

F

FREQUENCY

Frequency mengacu pada berapa kali pelanggan berinteraksi dengan produk atau melakukan transaksi dalam periode waktu tertentu. Frekuensi dijadikan ukuran karena dapat menunjukkan tingkat keterlibatan pelanggan dengan produk.

M

MONETARY

Monetary menunjukkan jumlah total yang dihabiskan oleh pelanggan untuk membeli produk dalam periode waktu tertentu.

R

RECENCY

Recency menunjukkan waktu interaksi terakhir pelanggan dengan produk.





Tahapan RFM Analisis

01

Membuat Model RFM

menetapkan skor recency, frequency, dan monetary value dengan menggunakan metode kuartil

02

Mengelompokan Pelanggan

Membagi database ke dalam beberapa kelompok berjenjang untuk tiap poin.

- 1 = Paling tinggi
- 2 = Tinggi
- 3 = Normal
- 4 = Rendah
- 5 = Paling rendah

03

Memilih Kelompok Yang Ditargetkan

Pilih segmentasi pelanggan dengan nilai tertinggi.

SEGMENTASI KELOMPOK RFM



01 CHAMPIONS

- Memiliki total Monetary Value yang paling tinggi

- Telah menyelesaikan transaksi
- Telah bertransaksi dalam kuantitas yang banyak

02 LOYAL CUSTOMER

- Memiliki total Monetary Value yang tinggi

- Telah bertransaksi dalam kuantitas yang relatif banyak
- Telah bertransaksi dalam waktu yang relatif baru dari transaksi terakhir

03 PROMISING CUSTOMER

- Memiliki total Monetary Value yang 'Baik'

- Telah bertransaksi dengan kuantitas yang normal
- Telah bertransaksi dalam waktu yang relatif lama dari transaksi terakhir

04 AT RISK

- Memiliki total Monetary Value yang sedikit

- Telah bertransaksi dengan kuantitas yang normal
- Telah bertransaksi dengan waktu yang relatif lama dari transaksi terakhir

SEGMENTASI KELOMPOK RFM

05 HIBERNATING

- Memiliki total Monetary Value yang sedikit
- Telah bertransaksi dengan kuantitas yang normal
- Telah bertransaksi dengan waktu yang lama dari transaksi terakhir

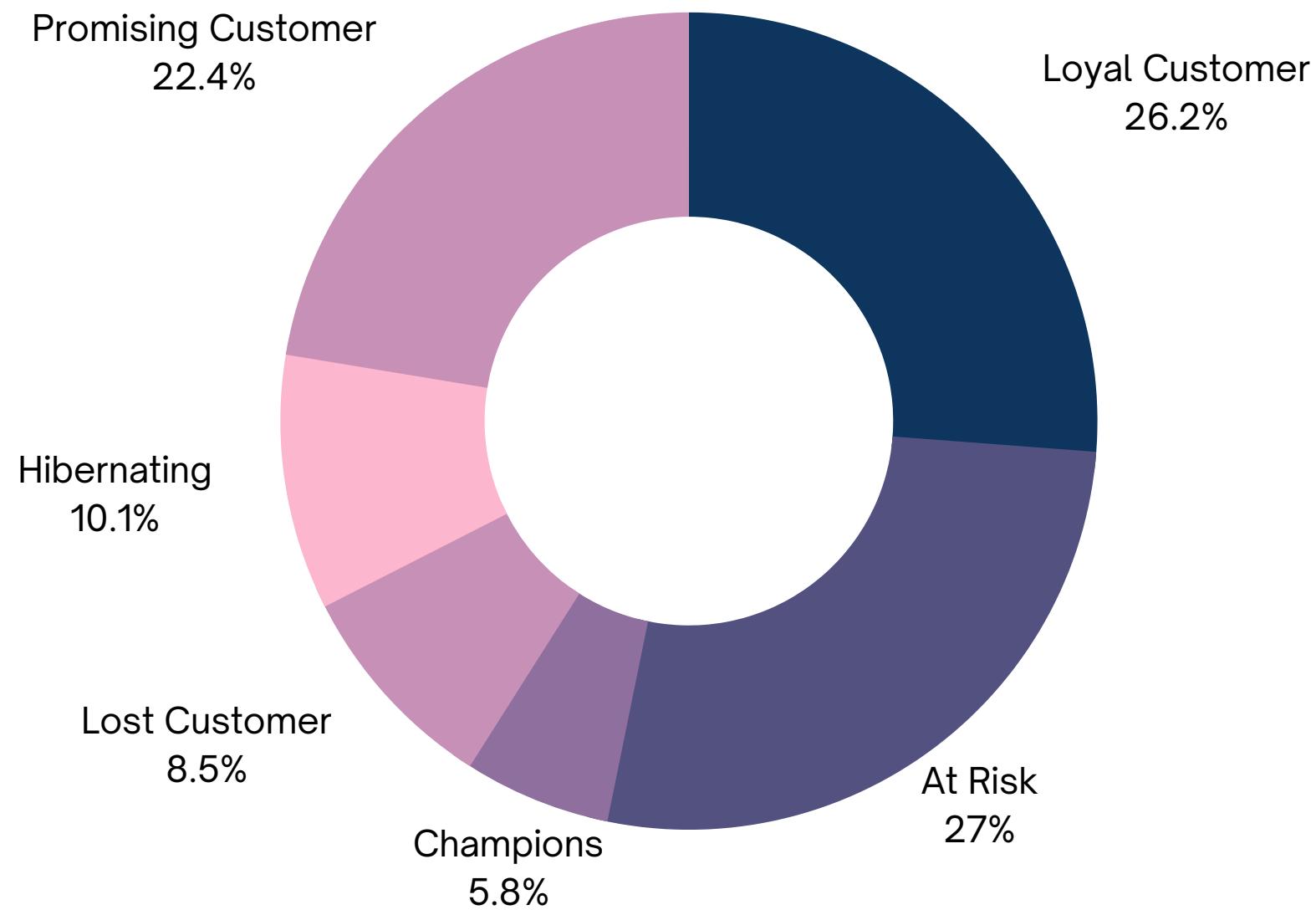
06 LOST CUSTOMER

- Memiliki total Monetary Value yang paling sedikit
- Telah bertransaksi dengan kuantitas yang paling sedikit
- Telah bertransaksi dengan waktu yang sangat lama dari transaksi terakhir



Composition

RFM Segmentation



Analysis:

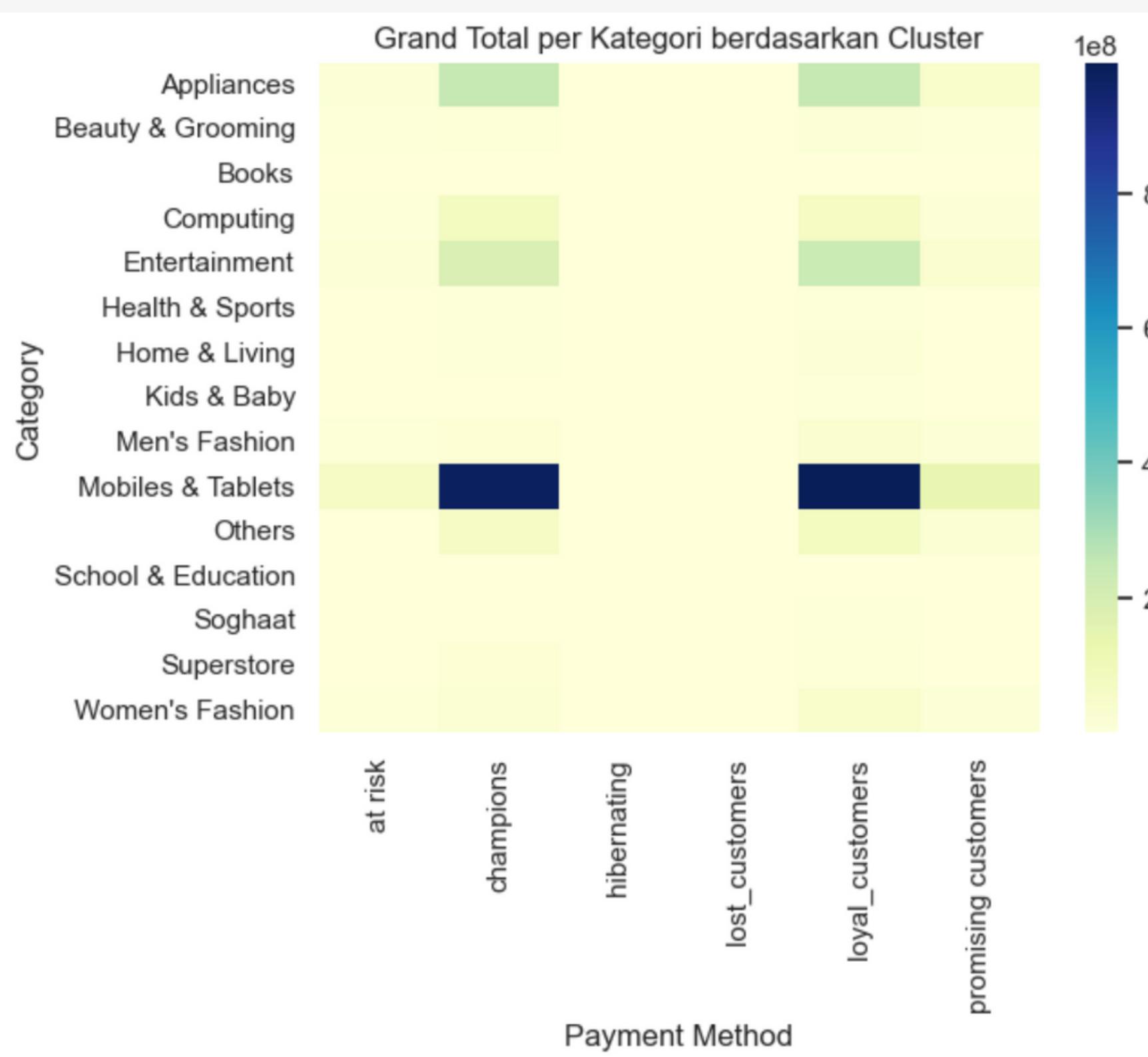
Urutan komposisi berdasarkan jenis segmentasi

1. Champions (ke-6)
2. Loyal Customer (ke-2)
3. Promising Customer (ke-3)
4. At Risk (ke-1)
5. Hibernating (ke-4)
6. Lost Customer (ke-5)

CLUSTER WISE : RFM SEGMENTATION



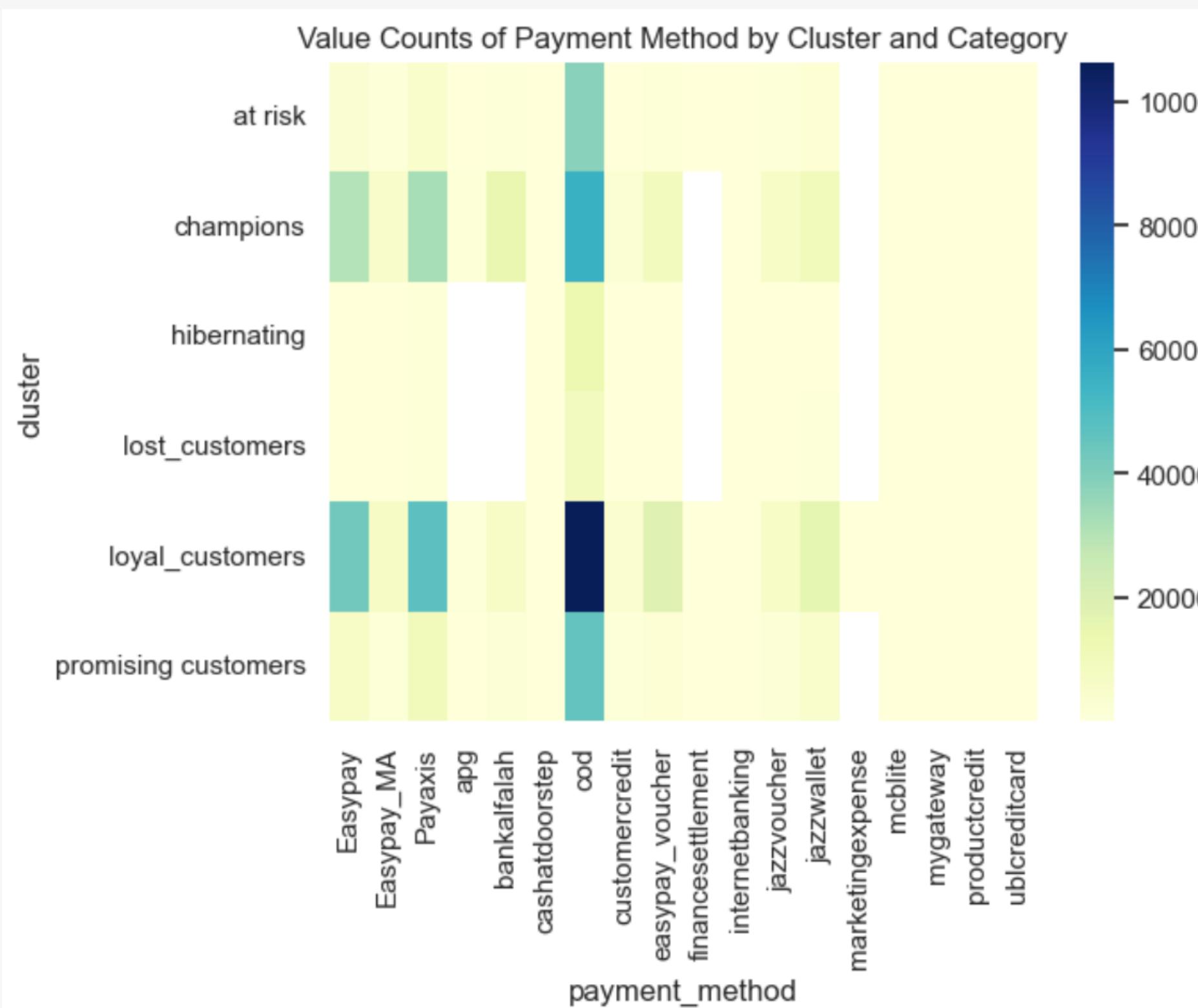
Spending Habits berdasarkan Segmen



ANALYSIS :

- Segmentasi "Champion" dan "Loyal Customer" memberikan kontribusi terbesar dalam grand total pada kategori Mobiles & Tablet. Terbesar kedua selanjutnya pada kategori Appliances

Payment Method berdasarkan Segmen



ANALYSIS :

- Penggunaan metode bayar COD merupakan pilihan favorit bagi segala jenis segmentasi
- Segmentasi "Champion" juga menjadikan COD sebagai pilihan utama pembayaran

Insight & Rekomendasi

STRATEGI	DESKRIPSI	CONTOH PENERAPAN
Customer baru (Product, Place, Pricing)	<p>Strategi komunikasi kepada customer baru secara online dengan penawaran produk atau layanan, yang berpotensi menghasilkan pendapatan.</p>	<ul style="list-style-type: none">• Bank - kerjasama melalui skema metode pembayaran untuk mendongkrak traffic
Customer Acquisition Initiatives	<p>Meningkatkan kemampuan situs untuk menghadirkan prospek baru secara berkelanjutan melalui berbagai teknik pemasaran online.</p>	<ul style="list-style-type: none">• SEO• PPC• Affiliate Marketing• Aggregators• Memperbaiki transaction flow (membantu meningkatkan conversion rate)

Insight & Rekomendasi

STRATEGI	DESKRIPSI	CONTOH PENERAPAN
Customer conversion dan Customer Experience	Investasi dalam fitur pelanggan baru di situs. Didasarkan pada kasus bisnis peningkatan tingkat konversi dan nilai pesanan rata-rata.	<ul style="list-style-type: none">• Kemampuan merchandising untuk menawarkan promosi yang disesuaikan• Sistem interaktif untuk membantu pemilihan produk• Panduan pembeli yang terdiri dari konten mendalam tentang produk atau multimedia (misalnya video yang menampilkan produk)

Insight & Rekomendasi

STRATEGI	DESKRIPSI	CONTOH PENERAPAN
Customer Development dan Growth Strategic	Investasi untuk meningkatkan pengalaman dan penyampaian penawaran kepada pelanggan yang sudah ada.	<ul style="list-style-type: none">• Personalisasi rekomendasi barang (upselling, cross selling)• Penggunaan E-mail marketing untuk berkomunikasi secara interaktif dan personal kepada customer• Mengikutsertakan komunitas untuk berpartisipasi dalam event Ecommerce

Thankyou

